interpretability characteristics functionalities trustworthiness materialisation straightforward organisational administration considerations interconnected administrative paperisation representation aforementioned reconstruction characteristic unrecognisable often-enforced accountability correspondence understandable communication informational philosophical understanding relationships international methodologies organisations possibilities consideration functionality transactional presentations characterised accessibility communicative intentionally investigation well-formed informationpresentation undercurrent perspectives requirements professional fundamentals organisation supplemented meaningful truthfulness information introduction architecture ecclesiastes particularly dependencies insufficient redefinition prerequisite environments re-inventors recognisable intelligence artificially consequently interweaving non-result requirement disposition disappeared transferred supervision linguistics maintaining utilisation well-formed meaningful components refuelling management activities understand archivists difference processing structured connection technology leeuwarder considered philosophy interwoven infosphere discipline connecting completely perception leeuwarden functional everything eventually interpares recognised additional individual connectors algorithms conceptual archivist thoroughly simplicity confronted migrations preferably interested documents standards behaviour therefore described different necessary technical concepts exchange semantic compound contexts supplier constant starting although examples partners received metadata recorded messages directly receiver internal archives consists somewhat abstract personal fixation research guidance entities delivers executed fuelling provided original objects because between message example content concept process display primary floridi element certain another station receipt answer result useful mostly almost simple factor cannot itself sender change system starts levels filing accept context principle integrity reveal boards expire filled results complex science petrol choose things parked begins vendor refill volume extras closed office visual occurs cotexo looked erased affect start given first second saving times above clear easily charts broker model there these about which other point three could place level found paper layer using their basic where event cases those sense still steps after makes types parts world often begin fixed field broad right debit total chain actor saved files essay quest grock issue basis years forms legal build terms while sites empty thing final reuse stock print again never study small click pages route reach rienk match grasp notes quick clash frans rules avoid annex daily fluid feith fruin union works asked leads drive speed train table extra handy euros stops gives price space bound latin views ideas holds focus meant least allow versa zeros going write video audio mouse email games disks refer major until arise drift added occur exist roman state quote time spans enter realm shown that this with from also time will must when same only used what part more pump have both they like some into make tank data made work kind form just user text each then even many fuel unit word been fact core such very need your find view long were come call want real look able hand high case hose most ways five file step give verb much rule eyes take 2010 2008 full meet sent stop side once card know soon main send role bank sees seem deep note link well 1961 born page sort rise turn lies base life lots maze easy 2013 show sure zone road stay park paid fill keep nine does scan hold vice ones copy down seen them ford 2015 lays open sets move 1948 puts 1960 code nice said 2009 true the has its all way his use was may own via say lot pay two ict due ten gas web gap she job act oil run of is in it pc

Archives in Liquid Times

Jaarboek 17

edited by

Frans Smit, Arnoud Glaudemans, Rienk Jonker

GOED VOOR
UW LOOPBAAN
BENEFICIAL FOR YOUR CAREER

De Hogeschool van Amsterdam | Archiefschool leidt archiefprofessionals op binnen het hoger beroepsonderwijs. Daarnaast verzorgen we na- en bijscholing voor de archiefsector.

HvA | Archiefschool:
▶ onafhankelijk, professioneel en praktijkgericht

Onze expertise:
▶ digitaal archief- en informatiebeheer
▶ duurzame toegankelijkheid
▶ persoonsgegevens en openbaarheid
▶ archiefrecht en auteursrecht

Maatwerk
Wat kan scholing voor uw organisatie betekenen? Op hva.nl/archiefschool vindt u het *complete cursusaanbod*. Neemt u gerust contact op voor een oriënterend gesprek.

Contact
📞 06 - 211 58 210
✉ archiefschool@hva.nl
🌐 hva.nl/archiefschool

HVA.NL/ARCHIEFSCHOOL

CREATING TOMORROW

Is Bill Gates part of your Swarm?

You are only six steps away from a connection to Bill Gates, co-founder of Microsoft. At least, that is what the "Six degrees of separation" theory tells us. In the digital world, those lines are even shorter. According to Facebook, its millions of users are connected by an average of 4,74 steps. Fact or fiction?

Fact is we are currently living in a 'swarm' and are all connected to each other. This requires a paradigm shift for the way we think about humankind, the work we do, and information storage.

In the book "De Zwerm. Een verhaal over de virtuele samenleving" ('The Swarm. A story about the virtual society') Doxis, in collaboration with three philosophers, explores the future of Dutch society. We point out three trends which will radically influence the work environment in the years to come: insecurity, atomisation and standardisation. What does this mean for the successful organising of both work and information? And what about the safeguarding of accessibility, quality control and sustainability?

These are important current issues, as proven by the many downloads, as well as a recent reprint, of our book! You can download the book 'De Zwerm' (in Dutch) at www.doxis.nl/dezwerm. Available as PDF and ePub.

The book 'De Zwerm' was created in collaboration with Filosofie Magazine.

DOXIS

No hassle information management: www.i-beez.nl
Structuring information, not people: www.somio.nl

When digital becomes the work of man: www.doxis.nl

MEMORIX MAIOR
Webbased Collection management

Books

Images

Maps

Genealogy

Video

Audio

Art

Audio

Herbarium

Archive

Papers

World Class Rapid Digitization
of Cultural Heritage

PICTURAE

More information: T +31 (0)72 - 53 20 444   contact@picturae.com   www.picturae.com

Archives in Liquid Times

Archives in Liquid Times

Archives in Liquid Times

edited by

Frans Smit
Arnoud Glaudemans
Rienk Jonker
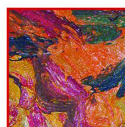
Foreword

Almost two decades ago, in an article in *The American Archivist*, I argued that research in archivistics (or: archival science) would save the archival profession, because research is the instrument for experimenting, inventing, changing, and improving – and a profession that is not involved in *"The endless cycle of idea and action, endless invention, endless experiment"* (T.S. Eliot) is doomed (Ketelaar, 2000). Often, archive professionals do not realize that many if not all managerial or practical questions can be solved more fundamentally when one allows for some theoretical and methodological reflection. "*Research,*" Barbara Craig (1996) wrote, *"cultivates a habit of examining received notions for their continuing pertinence and relevance."* (p. 110) Such a habit is essential for the archival professional who has to be equipped to deal with the constant change in his or her environment, effecting changes in records creation, preservation, communication, and use. As Arnoud Glaudemans and Jacco Verburgt declare in the first sentence of their essay in this volume: *"Any account of present-day archives should not only address practical, operational or managerial issues but also explicate the relevant theoretical issues regarding the specific nature and societal impact of digital information – if only because practical, operational or managerial issues, important as they obviously are, always presuppose some underlying theoretical framework."*

Archivistics offers such a theoretical framework, drawing on concepts like context, authenticity, findability, and access. In researching the ontological and epistemological archive(s), archivistics applies the archival method that is specific for the discipline, but it also adopts methods from other disciplines. This is evidenced by the various chapters in the recent book *Research in the Archival Multiverse* (Gilliland, McKemmish, Lau, 2016). But not only in methods: archivistics is increasingly profiting from what other disciplines can offer in conceptual and theoretical understanding of archival phenomena. So, for example, in performance studies dance may be understood as *"the choreographic activation of the dancer's body as an endlessly creative, transformational archive"* (Lepecki, 2010, p. 46). This resounds archivistics' concern with the fluidity of the archive as keeping former instantiations of a record 'in reserve', to be released not as exact copies but as re-enactments. And just as *"the originating instantiation"* of a dance keeps possibilities for later re-enactment in reserve, so gets each activation of a record along the records continuum extra significance in the light of subsequent activations.

Other 'archival turns' are also relevant to the theory, methodology and practice of archivistics. This volume shows what is brought to the archivistics' table from fields like media archaeology, speech act theory, information science, data science, philosophy, semiotics, genre studies, and organization science. At the same time, several essays in this volume indicate how archival theory and methodology can enrich other disciplines. In this way *Archives in Liquid Times* tries to cross disciplinary boundaries which so often keep scholarly and professional communities locked in their own discourse.

Eric Ketelaar

Literature

Craig, Barbara L. (1996). Serving the Truth: The Importance of Fostering Archives Research in Education Programmes, Including a Modest Proposal for Partnerships with the Workplace. *Archivaria 42*, 105-117

Gilliland, A.J., McKemmish, S., & Lau, A.J. (eds.) (2016). *Research in the Archival Multiverse*. Clayton, VIC: Monash University Publishing. http://dx.doi.org/10.26530/OAPEN_628143

Ketelaar, Eric (2000). Archivistics Research Saving the Profession. *American Archivist 63*, 322-340.

Lepecki, André (2010). The Body as Archive: Will to Re-Enact and the Afterlives of Dances. *Dance Research Journal 42* (2), 28-48.

Introduction

Archives are a reflection and a result of what happens in society. This means that they also (re)present society's changes and dynamics. Today, archives areundergoing fundamental changes in every aspect that one might think of. Digitisation and globalisation are turning our world upside down and reshape it. The same applies for archives, the archival profession and archival science. Therefore, in the entitling of this book, we decided to follow the metaphor of sociologist Zygmunt Bauman (2006), who characterized contemporary society as being in *"liquid"* times. By this he meant that present-day (western) society is in such a state of dynamics that it is difficult to get a grip on life. All foundations are shaking. In our opinion, Bauman has a case in stating that it is the main feature of the period we are now witnessing and are living in. That is why you are now reading a book that has the title: *"Archives in Liquid Times"*.

This book is inspired by several motivations and convictions. First, the editors are convinced that discussions and debates about archives in the digital age should become part of the broader discourse on information quality. This discourse should take place on several levels, for example on fundamental, conceptual and ethical issues. Our observation is that this integration is hardly happening. Archives and the archival community are in danger of being marginalised and 'doomed', when – and because of – losing connection to debates about for example the ethics of the internet and the development of data science. On the other hand, archival science's rich and detailed knowledge of the nature and function of records is hardly considered in fields like information science or philosophy. Building bridges between communities dealing with information quality is not a mere luxury – it is a necessity.

Our conviction is that paradigms and concepts that formed the basis of recordkeeping in the analogue world have lost their central place. Attempts to create a new paradigm or a new overall concept on archives in the digital information society have not yet been convincing. This reflects our liquid times, which the archival profession is also going through. The recent, extensive publication by Monash University tries to cover as much as possible research developments in the *"Archival Multiverse"* (Gilliland, McKemmish, Lau, 2016). In our view this multiverse *itself* is subject to radical changes regarding its own context, its subject matter, and its relevance to society.

Maybe we are all in the new landscape that Alessandro Baricco (2006) has described in his socio-cultural critique *"I barbari"*. In his account we are witnessing a mix in which all former boundaries between for example high and low culture and between fields of research fall apart. Most importantly he argues that present-day society is not interested in *"Why?"* questions anymore, but only in *"How?"* questions. His Barbarians surf their network all the time trying to find correlations without wondering about a reason or explanation of their environment. This network is essentially very liquid.

Several years ago, the editors of this volume concluded that, as professionals and experienced practitioners, they were getting a little lost. In their daily work, they could not derive enough grip and guidance from their own archival silo of concepts and methods anymore. They were also curious if these might be found elsewhere. Therefore, they decided to try to open their doors and look for new answers. They decided to make this journey in the unknown by trying to connect to the information philosophy of Luciano Floridi. The next step was decided upon during a lengthy discussion over some excellent Belgian beers at the ICA-congress in Brussels in 2013: we should produce a book on Information Philosophy and Archives. Our efforts have resulted in this publication. We hope it will be beneficial to academics, students, professionals and everyone else who is interested in disciplines like information philosophy, archival science, library science and data science. Its main emphasis however, still lies on the function and relevance of archives, and on how to keep and curate a necessary quality and accessibility of information – in between all other information professions in this digital age.

The contributions in this book are now summarised in the order in which they are published in this edition.

The first and second chapter are by Geert-Jan van Bussel. The first chapter is an overview of archival theories and their philosophical foundations, including modern digital diplomatics and the concept of the records continuum. In his second contribution Geert-Jan van Bussel presents a new theoretical framework for the archives in organisational context, based on a pragmatic approach. The "archive-as-is" is a part of Enterprise Information Management (EIM). In this framework the value of information, and the ensuing criteria for quality of records play a central part. The theoretical framework is positioned between modern diplomatics and the records continuum.

Rienk Jonker's essay is a theoretical exploration in which concepts of Luciano Floridi and concepts from archival theory are linked. It introduces an information model and a new definition of an information object. In this way a framework can be established that can be both of use to the archival professionals and community, as well as to disciplines like information philosophy and information theory.

In his contribution, Geoffrey Yeo concentrates on several theoretical perspectives, most notably on speech act theory (or philosophy). His essay considers how notions of 'information' might relate to a view of record-making and record-keeping that take account of speech act philosophy. It concludes that records have both social and informational roles. Speech act theory reminds us that records are not mere information objects or containers of facts, and it affirms that records do not simply dissolve into interpretation. At the point of inscription, a record and an action are interlinked: assertive, directive, commissive, or declarative.

In their article, Arnoud Glaudemans and Jacco Verburgt address the topic of today's archival transition from analogue to digital, by discussing and comparing Jacques Derrida and Vilém Flusser. Derrida stresses that, traditionally, an archive is largely defined by what he calls domiciliation, involving a hierarchical and centralized gathering and structuring of information. According to Flusser, the realm of digital,

algorithmically processed, information consists of what he calls technical images, which impose a shift from discursive (or textual) to dialogical (e.g., hyperlinked) information. This shift would make the traditional, centralized structure of the archive gradually obsolete, not from a Derridean 'deconstructivist' perspective, but from a techno-functionalist perspective. The discussion results in raising some theoretical and practical questions regarding the present-day archive, including the operational functionalities that need to be built into the digital for reasons of accountability.

The two following contributions are by Wolfgang Ernst. The first essay is inspired by Michel Foucault's *'L'Archéologie du Savoir'*. It explores media archaeology as a cross-disciplinary field of inquiry, that consists of a radically material and mathematical approach to the study of cultural change, memory, and knowledge tradition, and even the very category of time itself. The second essay concentrates on audio-visual information. Archives, today, can be re-defined in terms of negentropic systems. How can not only material traces and textual documents, but temporal expressions (or movements) themselves be preserved for future historiographies? Ernst's answer lies in discovering, reflecting and techno-mathematically realising new options of flexible access.

Fiorella Foscarini and Juan Ilerbaig reflect on the basic concept of context. They use a semiotic approach in which they provide insights that point to an expanded and more dynamic view of text-context relationships. Rhetorical Genre Studies (RGS) offer a set of concepts and analytical tools that shed light on the social context of records creation and use. By looking at intertextual relationships in the archives, archivists can develop a better insight as to the mechanisms involved in the choices made by record creators and users; an insight that in turn elucidates context as a situated construct.

The following chapter is a reflection by Charles Jeurgens on the position of recordkeeping in the digital age, and on accountability and transparency in view of the current data-flood. He argues that the present and mainstream views of appraisal in the recordkeeping community should radically change. We should focus on understanding and managing the assemblages between data and the processing mechanisms (for instance algorithms) in situated practices.

Anne Gilliland's essay is about metadata. It puts the concept of metadata in historical perspective. In the past decades the concept has had a profound influence on archival theory. The essay raises fundamental questions about the relationship between records and metadata, about metadata practices and standards and about their ethical implications.

Another basic concept in archival theory: provenance is the subject of the essay of Giovanni Michetti. Provenance in the archival domain has moved from a simplistic one-to-one relationship to a multi-dimensional approach. It is now being understood as a network of relationships between objects, agents and functions. Any lack of control over provenance determines some uncertainty which in turn affects trust in digital objects, so we will have to develop new ways to approach and ascertain digital provenance.

Frans Smit reflects on another basic concept in archival theory: authenticity. His essay gives an overview of how this concept is used regarding archives. He argues that to gain a better understanding of the authenticity of records in a digital environment, it is necessary to redefine the nature of records and their context. He uses the concept of hyperobjects, originating from ecological philosopher Timothy Morton, to gain a better understanding of records in a data-immersed society, and as a starting point to rethink the way authenticity of records in such an environment can be asserted.

Information ethics is the central issue of the essay by Martijn van Otterlo. He explores the ethics concerning digital archives from the perspective of data science, and with an emphasis on the role of algorithms. Ethical principles, about access, have been formalised and communicated in the form of ethical codes, or: codes of conduct. This last topic brings us from the intended, human archivist in physical domains to the intentional, algorithmic archivist, of: *algivist*, in the digital domain. Which codes of conduct should be made for the latter, and how to implement them?

The book concludes with interviews in which two internationally renowned scholars. Archival theorist Eric Ketelaar and information philosopher Luciano Floridi share their reflections on the subjects raised in this book. The interviews mainly concern the nature (and future) of records, the (digital) ethics concerning archives, and the role that the various professional communities on information should play nowadays.

As editors we hope that this book will stimulate the exchange of ideas, concepts and critical thinking from different areas. We also hope that it can be of help in taking further steps in building bridges between archival thinking and the many other fields of research concerning the quality of information. We hope that the book offers some anchors of thought in these liquid times, maybe even anchors for new programs of research into the nature, the position and the societal importance of archives in our data-immersed digital information society.

Acknowledgements

occasions and we thank them for their hospitality. We are grateful to Prof. Charles Jeurgens and the University of Amsterdam for giving the opportunity to present our book during a special symposium. And finally, we would like to thank Hans Berende, and with him the Stichting Archiefpublicaties, for the possibility to produce and publish this book. If it were not for Hans and S@P, this book would never have come into existence.

Arnoud Glaudemans
Rienk Jonker
Frans Smit

Literature

Baricco, A. (2006). *I barbari, saggio sulla mutazione*. Milano: Feltrinelli.

Bauman, Z. (2007). *Liquid Times, Living in an Age of Uncertainty*. Cambridge (UK): Polity Press.

Gilliland, A.J., McKemmish, S., & Lau, A.J. (eds.) (2016). *Research in the Archival Multiverse*. Clayton, VIC: Monash University Publishing. http://dx.doi.org/10.26530/OAPEN_628143

Glaudemans, A., Jonker R., Smit F. (2015). The right people and the right forces need to get together, Interview with Prof. Luciano Floridi, *Archievenblad, L-F, 01/2015*. Retrieved from http://www.youblisher.com/p/1048348-Archievenblad-uitgave-van-de-Koninklijke-Vereniging-van-Archivarissen-in-Nederland/

Table of Contents

The Theoretical Framework for the 'Archive-As-Is'. An Organization Oriented View on Archives

Part I. Setting the Stage: Enterprise Information Management and Archival Theories

1. Introduction

1.1. The problem: Information Chaos within organizations

The definition of a business strategy is a common practice to capitalize on new market opportunities and to do better than direct competitors. Projected on the information management processes of an organization, a business strategy clarifies how information can be used for reaching business objectives (Baets, 1992; Peppard and Ward, 2016). It can be used for quick responses to needs of customers, adjustments to changes in the organizational environment, and improvements in competitiveness. Most of this information is recorded in different types of information objects that are embedded within organizational business processes and are, as such, important business assets.

Enterprise Information Management (EIM) tries to enable organizations to secure these business assets across the complex landscapes of organizational departments, legacy systems, corporate and regulatory policies, business content, and big data (Chaki, 2015). It organizes the information value chain in capturing, structuring, describing, preserving, and governing information objects across organizational, temporal, and technological boundaries to allow business strategies to reach their objectives (Van Bussel, 2012abc; Van de Pas and Van Bussel, 2015ab). It has not been overly successful, because it concentrates almost exclusively on structured information (objects), as the result of being influenced extensively by computer science. But more than 80 % of all information objects in organizations are unstructured and with big data on the rise, that amount is growing quickly (Van Bussel, 2012b).

More than forty years ago, Alvin Toffler (1970) coined the term 'information overload'. Today's world is characterized by an increasing information flood, completely fulfilling Toffler's forecasts. According to IDC, in 2020 the digital universe (the information objects created and copied annually) will reach 44 zettabytes (44 trillion gigabytes) (Turner et al, 2014). Because EIM has neglected the management of unstructured information objects, many of these objects

cannot be quickly found when needed. Knowledge workers spend up to 40% of their working day searching for information (objects) (Nel, 2015; Naidoo and Nsibirwa, 2015). They spend 15-25% of their time on information-related tasks (Brocke et al, 2011). An 'information chaos' caused by the unability of EIM to capture this large influx of unstructured information objects compromises the ability of organizations to reach business objectives. This chaos is the rule rather than the exception in contemporary organizations (Redman, 2004).

The abundance of (structured *and* unstructured) information objects leads to organizational challenges. To facilitate fail-proof information management guaranteeing accountability, compliance, and security is by no means new (Hausmann et al, 2014; Patnayakuni and Patnayakuni, 2014). Until a few years ago, organizations captured and controlled information objects in an infrastructure that did not cross the borders of the organizational structure. If accountability, compliance, security, or other business-related issues arose, there was a single 'point of control' defined (Davenport and Prusak, 1997). That 'point of control' became diffused with the ongoing integration of business processes between different organizations, stimulated by sharing information objects through (for instance) social media (McAfee, 2006) and the breakthrough of supply chain and ERP systems causing information integration (Srinivasan and Dey, 2014). As it became common practice to share information objects between different parties, it could become difficult to ascertain which of the integrated process owners was responsible for accountability, compliance, security, or information accessibility. It is proving challenging for traditional ways, methods and technologies to achieve the expected information quality, compliance and information governance (Van de Pas and Van Bussel, 2015ab). Guaranteeing an accountable, compliant, transparent, and effectively performing organization in a dynamically changing ICT environment, recognizing both structured and unstructured information objects, is problematic. EIM's focus is changing to incorporate unstructured information objects, but lacks the theoretical foundation to do so effectively.

1.2. The solution: the organizational archive and its records?

The key for such a theoretical foundation for EIM may be 'the archive' (Van de Pas et al, 2016). For defining business strategies, Smith and Steadman (1981) already acknowledged organizational archives as crucial resources. They are very important for organizational accountability, business process performance, and reaching business objectives. They have, unfortunately, not been recognized as such for many years and for that reason have been badly managed by organizations, do not meet quite common quality requirements, and are almost non-contextual (Redman, 2004; Groth, 2007). Without these characteristics, it is impossible to realize the primary goals of archives: a reliable reconstruction of past happenings, delivering evidence, and meaningful production (Van Bussel, 2012abc), extremely diminishing their organizational value.

Organization-wide management of archives has not been a common functionality for EIM (Serova, 2012). The neglect in the management of organizational archives has resulted in [1] fragmented storage of both structured and unstructured information objects in a variety of information systems, unconnected with their metadata and the organizational archive they belong to; [2] fragmented metadata,

separated from the information objects that caused their genesis and not embedded into the metadata layers of the organizational archive, leading to a loss of contextuality; and [3] a declining quality of information objects, because their provenance, integrity, and preservation are in peril (Van Bussel, 2016).

Two concepts are essential for integrating structured *and* unstructured information objects within EIM to exploit the value(s) of information in defining effective business strategies: records and archives. Records are combinations of information objects (structured and unstructured data, data sets, and data objects) and their metadata, generated and used in the course of (business) processes, actions, and transactions, stored in an organizational (or personal) archive, irrespective of format used, with a unique (fixed or reconstructable) content, context, and structure, and retained and preserved for whatever reason organizations (or individuals, groups, or families) want to set them aside (business use, compliance, accountability, evidence, future reference, curiosity, historical value, extension of human memory, etc.) or for whatever period of time they (or parts of them) are retained (Van Bussel, 2016; Yeo, 2007). Archives (or data stores) are organizational or personal constructs, embedded in and enriched by metadata about their creation, organizational environment, and management, in which records (from the moment of their creation) are persistently stored and managed with the objectives of reliably reconstructing the past, delivering evidence, and realizing meaningful production.[1] The term can be used for *any* construct of records that is meant to be retained, like YouTube, Twitter, Pinterest, etc., but also more traditional organizational or personal compositions of records (Van Bussel, 2012b). Both concepts do not differentiate between structured and unstructured information objects.

To allow for the integration of structured and unstructured information objects, EIM needs a theoretical foundation based on records and archives that is aimed at realizing organizational objectives.

1.3. The objective: a theoretical foundation

Both computer and Information science cannot be expected to define this theoretical foundation for EIM, although they have developed many useful concepts and theories.

As shown in Tables 1 and 2, an analysis of the contents and abstracts of five top journals each for computer and information science from 2010-2016 shows that both sciences do not really acknowledge the concepts records and archives. They are rarely used, even while there are many articles in these journals describing information objects within business processes used for organizational objectives that are traditionally known as records or archives. In these journals, they are called digital artefacts, documents, data objects, repositories, archival collections, archival documents, or storage platforms. These articles were *not* included in the analysis visualized in Table 1 and 2, just like the three articles using the terms 'archiving' and 'archivists'. In the end, only 25 articles (from the 5.319 articles reviewed) mention the concept records or archive(s) (or both) in its title or abstract.

---

[1] In this interpretation of the concept 'archive', I am following the Dutch archival tradition that uses the term 'archive' to designate an organizational (or personal) construct of [1] current (or active) records; [2] semi-active or semi-current records; [3] inactive or non-current records; and [4] permanent records, the whole body of records of continuing value of an organization or person.

Twelve articles use the concepts only to indicate specific information objects ('(personal) health records', 'medical records', 'patient records', 'personal records', 'archives'). Eleven articles use the concepts to indicate aspects of the management of records and archives: 'records management', 'records management systems', 'records management metadata', 'records laws', 'archival finding aids', 'records management strategies', 'record search', and '(functional) records classification'. One article explores the relationship between information culture and records management and offers an interesting theoretical discourse, but it is not meant to be a theoretical framework for records and archives (Sundqvist and Svärd, 2016). In the articles analysed, there is only one that offers a theoretical framework, a formal model for digital achives as cultural heritage (Ferro and Silvello, 2013). It is not a full-scale framework for records and archives, but a very useful application of digital library research to archives as cultural heritage. Ferro and Silvello propose a formal model, called NEsted SeTs for Object hieRarchies (NESTOR). The model is used to extend the 5S model, a unified formal theory for Digital Libraries. It allows for a definition of a digital archive as a specific digital library able to cope with the peculiar features of archives (as context and hierarchy) and provides archives with the full wealth of digital library technologies and methods.

Journal	Impact	Articles (Total)	Articles (Concepts mentioned)	Articles (frameworks defined)
Computer Science				
MIS Quarterly: Management Information Systems	6.984	358	1 (record)	0
Information Systems Research	4.397	408	3 (record)	0
ACM Computing Surveys	3.405	417	0	0
Journal of Management Information Systems	3.036	306	0	0
IEEE Transactions of Industrial Informatics	2.513	1.086	0	0
Information Science				
Library and Information Science Research	1.629	272	3 (record)	0
Journal of the Association for Information Science and Technology	1.601	1.340	5 (record, archival finding aid)	0
Information and Organization	1.306	90	1 (record)	0
International Journal of Information Management	1.173	541	9 (record, archives)	0
Information Processing and Management	0.897	501	3 (record, archives)	1
	5.319	**25**		**1**
	100%	**0,47 %**		**0,0188 %**

Table 1. Analysis of content of applicable top journals 2010-2016 (Scimago Journal and Country Rank (April 26, 2017))

Journal	Article
MIS Quarterly: Management Information Systems	1. Kohli, R., and S.S.L. Tan (2016). 'Electronic Health Records: How Can IS Researchers Contribute to Transforming Healthcare?', *Mis Quarterly*, Vol. 40, No. 3, pp. 553-573.
Information Systems Reseach	2. Ozdemir, Z., J. Barron, and S. Bandyopadhyay (2011). 'An analysis of the adoption of digital health records under switching costs', *Information Systems Research*, Vol. 22, No. 3, pp. 491-503.
	3. Oborn, E., M. Barrett, M., and E. Davidson (2011). 'Unity in diversity: electronic patient record use in multidisciplinary practice', *Information Systems Research*, Vol. 22, No. 3, pp. 547-564.
	4. Mishra, A.N., C. Anderson, C.M. Angst, and R. Agarwal (2012). 'Electronic health records assimilation and physician identity evolution: An identity theory perspective', *Information Systems Research*, No. 3 (part 1), pp. 738-760.
Library and Information Science Research	5. Kettunen, K., and P. Henttonen (2010). 'Missing in action? Content of records management metadata in real life', *Library & information science research*, Vol. 32, No. 1, pp. 43-52.
	6. Sinn, D., S.Y. Syn, and S.M. Kim (2011). 'Personal records on the web: Who's in charge of archiving, Hotmail or archivists?', *Library & Information Science Research*, No. 4, pp. 320-330.
	7. Oltmann, S.M., E.J. Knox, C. Peterson, and S. Musgrave (2015). 'Using open records laws for research purposes', *Library & Information Science Research*, Vol. 37, No. 4, pp. 323-328.
Journal of the Association for Information Science and Technology	8. Nov, O., and W. Schecter (2012). 'Dispositional resistance to change and hospital physicians' use of electronic medical records: A multidimensional perspective', *Journal of the Association for Information Science and Technology*, Vol. 63, No. 4, pp. 648-656.
	9. Steele, R., K. Min, K., and A. Lo (2012). 'Personal health record architectures: technology infrastructure implications and dependencies', *Journal of the American Society for Information Science and Technology*, Vol. 63, No. 6, pp. 1079-1091.
	10. Li, T., and T. Slee (2014). 'The effects of information privacy concerns on digitizing personal health records', *Journal of the Association for Information Science and Technology*, Vol. 65, No. 8, pp. 1541-1554.
	11. Huvila, I., Å. Cajander, M. Daniels, and R.M. Åhlfeldt (2015). 'Patients' perceptions of their medical records from different subject positions', *Journal of the Association for Information Science and Technology*, Vol. 66, No. 12, pp. 2456-2470.
	12. Freund, L., E.G. Toms (2015). 'Interacting with archival finding aids', *Journal of the Association for Information Science and Technology*, Vol. 67, No. 4, pp. 994-1008.
Information & Organization	13. Davidson, E. J., C.S. Østerlund, and M.G. Flaherty (2015). 'Drift and shift in the organizing vision career for personal health records: An investigation of innovation discourse dynamics', *Information and Organization*, Vol. 25, No. 4, pp. 191-221.
International Journal of Information management	14. Külcü, Ö., and T. Çakmak (2010). 'Evaluation of the ERM application in Turkey within the framework of InterPARES Project', *International Journal of Information Management*, Vol. 30, No. 3, pp. 199-211.
	15. Xie, S.L. (2013). 'National strategy for digital records: Comparing the approaches of Canada and China', *International Journal of Information Management*, Vol. 33, No. 4, pp. 697-701.
	16. Shaw, N. (2014). 'The role of the professional association: a grounded theory study of electronic medical records usage in Ontario, Canada', *International Journal of Information Management*, Vol. 34, No. 2, pp. 200-209.
	17. Asma'Mokhtar, U., and Z.M. Yusof (2015). 'The requirement for developing functional records classification', *International Journal of Information Management*, Vol. 35, No. 4, pp. 403-407.
	18. Vilar, P., and A. auperl (2015). 'Archives, quo vadis et cum quibus?: Archivists' self-perceptions and perceptions of users of contemporary archives', *International Journal of Information Management*, Vol. 35, No. 5, pp. 551-560.

International Journal of Information management	19. Sundqvist, A., and P. Svärd (2016). 'Information culture and records management: a suitable match? Conceptualizations of information culture and their application on records management', International Journal of Information Management, Vol. 36, No. 1, pp. 9-15.
	20. Gagnon, M.P., D. Simonyan, E.K. Ghandour, G. Godin, M. Labrecque, M. Ouimet, and M. Rousseau (2016). 'Factors influencing electronic health record adoption by physicians: A multilevel analysis', International Journal of Information Management, Vol. 23, No. 3, pp. 258-270.
	21. Asma'Mokhtar, U., Z.M. Yusof, K. Ahmad, and D.I. Jambari (2016). 'Development of function-based classification model for electronic records', International Journal of Information Management, Vol. 36, No. 4, pp. 626-634.
	22. Mokhtar, U.A., and Z.M. Yusof (2016). 'Records management practice: The issues and models for classification', *International Journal of Information Management*, Vol. 36, No. 6, pp. 1265-1273.
Information processing and Management	23. Romero, F.P., I. Caballero, J. Serrano-Guerrero, and J.A. Olivas (2012). 'An approach to web-based personal health records filtering using fuzzy prototypes and data quality criteria', Information Processing & Management, Vol. 48, No. 3, pp. 451-466.
	24. Ferro, N., and G. Silvello (2013). 'NESTOR: A formal model for digital archives', Information Processing & Management, Vol. 49, No. 6, pp. 1206-1240.
	25. Amini, I., D. Martinez, X. Li, and M. Sanderson (2016). 'Improving patient record search: A meta-data based approach', Information Processing & Management, Vol. 52, No. 2, pp. 258-272.

Table 2. Articles mentioning records and/or archives in applicable top journals 2010-2016

The professional practice of records management (or recordkeeping), recognizing both concepts, aims to support organizations in their business conduct and should be aligned with business and information systems, risk management, and information governance (McLeod and Lomas, 2015). Although it should have been aligned with EIM, in organizational reality it is not (Alalwan and Weistroffer, 2012). Most organizations do not align records management with business processes and strategies (Van Bussel, 2016). Records Management is based on best practices, pragmatic considerations and borrowed theories from other disciplines such as archival science, information science and management. It lacks its own theoretical basis, as an analysis of the contents and abstracts of the 392 primary articles of its only scholarly journal, Records Management Journal (Impactscore: 0.324) shows. The two articles defining encompassing frameworks for records management are based directly on theories from archival science (Yusof and Chell, 2002; Ismail and Jamaludin, 2009). Its reputation is that of 'the handmaiden of archives administration' (McLeod and Lomas, 2015, p. 349), a keeper of 'old documents', and an inconvenience or technicality. It is one of the main reasons for the organizational misunderstanding about the value of records management.

For EIM to find a theoretical foundation based on records and archives, only archival science seems to offer applicable, encompassing theoretical frameworks. There are two different views within archival science, the Records Continuum Theory and Digital Diplomatics. The theories focus on the cultural and evidential value of archives respectively. They do not pay much attention to the (organizational

or personal) construction of archives and their value for reaching organizational objectives and defining business strategies. It is remarkable, for instance, that in the more than 1.000 pages of *Research in the Archival Multiverse*, the most recent collection of essays on archival science research (Gilliland, McKemmish and Lau, 2016), the organizational (or personal) construction of archives receives no attention at all. The analysis of these two theoretical frameworks will be very important in defining a new one, more aimed at organizational value, reaching business objectives and defining business strategies. This new framework could be the theoretical foundation needed for EIM to use records and archives for reaching business objectives and in defining and realizing business strategies. Defining this new theoretical framework is the objective of the second part of this article.

1.4. Research methodology

This new theoretical framework is a result of my long-term research into the relationships between organizational accountability, digital archiving and EIM (2008-2016). During this research, an interpretive research approach was followed, primarily based on Orlikowski and Baroudi (1991). In this approach, in order to explore phenomena without imposing an *a priori* understanding, a non-deterministic perspective is necessary. The research for this article is based on:

1. An analysis of a corpus of scientific literature, based on the literature review methodology of Okoli and Shabram (2010). This corpus consisted of 1152 conference papers, journal articles, working papers and books. These items were collected using key word search in the Digital Library of the University of Amsterdam, IEEE Xplore Digital Library, ACM Digital Library, Google Scholar, Microsoft Academic Search, EBSCO, Emerald Insight, and Paperity.
2. An analysis of the application descriptions of 17 case studies, organized and analysed for the research reported in Van Bussel and Ector (2009) about digital archiving, organizational accountability and governance in public organizations. These case studies were organized according to the case study methodology, designed by Yin (2003), supplemented with Benbasat, Goldstein and Mead (1987) for their method of action research.
3. In-depth, semi-structured interviews with ten records management, EIM, and business specialists about the analysis of the corpus of literature, the case studies, and the new theoretical framework in its different stages of design.

1.5. Outline of this article

This article is published in two parts. This first part is setting the stage. In the introduction, EIM and its lack of a suitable theoretical foundation is introduced. This is followed with a delineation about the archival renaissance in the last decade of the twentieth century, when the 'archive' was 'reborn' as a 'theoretical archive', almost completely dissociated from organizational practice and characterized as a conceptual domain for many disciplines. After this, the leading archival theories in the first decades of the twenty-first century are discussed. The first part ends with a conclusion about the value of these archival theories for the theoretical framework that is discussed in the second part of the article. This second part will be an in-depth discussion of the new framework, developed as a way for EIM to use records and archives for reaching business objectives.

2. The Archival Renaissance

2.1. Foucault and Derrida

Since the early 1990s, in the wake of a new edition of Michel Foucault's (1992) *L'Archéologie du savoir*, archives have become the conceptual domain of a range of disciplines, most notably literary and cultural studies, philosophy, and anthropology. Foucault was, in essence, the pioneer of 'the theoretical archive' that is entirely dissociated from its conventional definition(s) and practices. The Foucauldian archive does not reproduce but produces meaning; it is not a monument for future memory, but a 'document' for possible use (Foucault, 1975, p. 193). Jacques Derrida, who reformulated the notion of an archive in terms of psychoanalysis, has pointed out in his highly complex 'Mal d'Archive' (1995a, p. 141) that 'rien n'est moins sûr, rien n'est moins clair aujourd'hui que le mot d'archive' ('nothing is less reliable, nothing is less clear today than the word 'archive'': Derrida, 1995b, p. 57). For Derrida (1995a, p. 34) the process of archivization (a term which meaning is not always clear) 'produit autant qu'elle enregistre l'événement' ('produces as much as it records the event': Derrida 1995b, p. 17).[2]

We are confronted with what Marlene Manoff (2004, p. 14) has called 'the postmodern suspicion of the historical record'.[3] For archives are not passive receptacles: they shape and control the way the past is read. As Derrida (1995a, p. 15-16 (note 1); 1995b, p. 10-11 (note 1)) says, there is no power without control of 'the archive'. But, at the same time, 'postmodernists' are ambivalent about archives. They doubt the dominance of historical narratives (and that is not without reason). They view archives as 'traces of missing or destroyed universes of records and activity' and as 'trick mirrors distorting facts and past realities in favour of the narrative purpose' of authors and audiences (Cook, 2001, p. 9). Nevertheless, they resort to history and historical analyses. Foucault's historical studies on criminology and sexuality are exemplary examples (Foucault, 1975, 1976, 1984).

2.2. An inflation of terms

Archives are 'loosening and exploding' (Manoff, 2004. p. 10). In the resulting inflation of the term, archives have become 'loose signifiers for a disparate set of concepts' (Manoff, 2004, p. 10), such as: the 'social archive' (Greetham, 1999), the 'raw archive' (Galin and Latchaw, 2001), the 'postcolonial archive' (Shetty and Bellamy, 2000), 'the popular archive' (Lynch, 1999), 'the ethnographic archive' (Marcus, 1998), 'the geographical archive' (Withers, 2002), and 'the liberal archive' (Joyce, 1999). It leads Marta Voss and Paul Werner (1999) to dwell on 'the poetics of the archive'. It has been suggested that the changes in information technology are responsible for this inflation. The technological revolution, after all, has altered 'our relationship to the archive' (Voss and Werner, 1999, p. ii), it

---

[2] I will not elaborate here on the poststructuralist view of the archive, as expressed by Foucault and Derrida. For introductory reading: G. Bennington, 'Derrida's Archive', *Theory, Culture & Society*, Vol. 3 (2014), No. 7/8, pp. 111-119; B. Brothman, 'Declining Derrida: integrity, tensegrity, and the preservation of archives from deconstruction', *Archivaria*, Vol. 48 (1999), Fall, pp. 64-89; K.O. Eliasson, 'The Archives of Michel Foucault', E. Røssaak (ed.), *The Archive in Motion: New Conceptions of the Archive in Contemporary Thought and New Media Practices*, Oslo, Novus Press, 2010, pp. 29-51; S. Lubar, 'Information culture and the archival record', *The American Archivist*, Vol. 62 (1999), Spring, pp. 10-22; M. Morris, 'Archiving Derrida', *Educational Philosophy and Theory*, Vol. 35 (2003), No. 3, pp. 297-312; and R. Vosloo, 'Archiving otherwise. Some remarks on memory and historical responsibility', *Studia Historiae Ecclesiasticae*, Vol. 31 (2005), No. 2, pp. 379-399.

changed 'the archive' into 'a metaphor for what we are not yet able to grasp about the nature of digital collections' (Manoff, 2004, p. 10), and it resulted in such an addiction to live connections to cyberspace that to lose them is 'to die', that is 'to be disconnected from access to the archives, not jacked-in or not in real time' (Mackenzie 1997, p. 66). Andreas Huyssen (2000, p. 33) and Marlene Manoff (2001, p. 371-372) argue that the development of information technology has led to anxiety about the preservation of cultural heritage, to fears about the loss of historical awareness resulting from a loss of roots in time and space, and to cultural and historical amnesia because of information technology defects. Both argue that technological changes have bolstered an obsession with historical information. That is possible, just as it is undeniable that information technology changes affect information growth and influence the way organizations create, use, and store information (Van Bussel, 2012a). But it is, in my opinion, doubtful if they *caused* the inflation of the *term* 'archive'. The continuous use of that *term* in multidisciplinary contexts for very different types and collections of information objects and records seems a more probable cause for that inflation.

2.3. The 'Archival Turn'

The terms 'archive' and 'archives' seem to be used as keywords for questions of, among others, memory, evidence, taxonomy, governance, and justice. This preoccupation with 'the archive' is characterized as the 'archival turn', which can be seen as a follow-up (or a part) of the 'historical turn' (McDonald, 1996). The term signifies the repositioning of 'the archive' as a subject of investigation, more than as a mere site for research or a collection of records for research use. As Ann Stoler (2002, p. 87) states, using poststructuralist arguments: the 'archival turn' means looking to archives more as epistemological experiments of the past than as historical sources, as cross-sections of contested knowledge, as transparencies inscribed with power relations, and technologies of rule. The 'archival turn' positions 'the archive' as, as Jacques Derrida (1995a, p. 60) states, 'n'est pas la question d'un concept dont nous disposerions ou ne disposerions pas déjà au sujet du passé, …. C'est une question d'avenir' ('[not] the question of a concept dealing with the past which already might be at our disposal or not at our disposal, …[but rather] a question of the future': Derrida 1995a, p. 27). It is an intriguing concept that opened doors for exhilarating research. This 'turn' has stimulated scientists to research the role of 'the archive' in social conditions and in postcolonial, post-trauma, and post-conflict societies. Seen as 'the decolonisation of the archive', it is situated in discourses on postcolonialism and postcoloniality (Stoler, 2002). It is studied as a political space, as a societal concept for the promotion of power, nationalism, surveillance, and for the silencing of alternative narratives (Burton, 2005; Chakrabarty, 2000; Faulkhead, 2009; Ketelaar, 2002; Stoler, 2009; McKemmish et al, 2011). 'The archive' is used as a concept in themes as race and

---

[3] Postmodernism is used as a rather loose label to identify a number of theoretical approaches developed since the 1960s. Poststructuralism, as a much more precise but less inclusive term, is used to refer to the French theorists Jacques Derrida, Michel Foucault, Jacques Lacan, Julia Kristeva, and Roland Barthes. They demonstrate the dependence of structures on what they try to eliminate from their systems. They diverge from one another in many ways, but they have in common the attempt to uncover the unquestioned dependencies and methaphors that uphold social and cultural norms. Postmodernism also includes theorists that are influenced by but are not within poststructuralism: Gilles Deleuze, Felix Guattari, Jean-Francois Lyotard, and Jean Baudrillard. Many theorists (like Gatyatri Spivak, Judith Butler, and Donna Haraway) are critical of postmodern theory but find elements of it very useful. Postmodernists and poststructuralists do *not* constitute a single school and there is as much disagreement among them as between them and other types of theory.

ethnicity, identity, gender, sexual orientation, and transnational approaches of migration (Kaplan, 2000; Cvetkovich, 2003; Wurl, 2005; Dunbar, 2006; Rawson, 2009; Campt, 2012; White, 2017). It is about the epistemological and symbolic role of 'the archive' in a trans-disciplinary, multicultural, pluralistic, and increasingly interconnected and globalised world (Dunbar, 2006; Kaplan, 2000; Wurl, 2005). The organizations that generated the archives disappeared from consideration.

2.4. Rethinking the archive

Archival scholars, as Upward (1996, 1997), Brothman (1999, 2001), Cook (1997, 2001), Ketelaar (1999, 2000a, 2017), Nesmith (1999, 2002), and, recently, Wood et al (2014) are engaged in re-thinking and debating archival theory in the wake of the 'archival turn', using postmodern (especially poststructuralist), structurationist, and postcustodial theories[4] and the concept of (Derridean) archiviology, 'une science générale de l'archive, de tout ce qui peut arriver à l'économie de la mémoire et à ses supports, traces, documents …' (Derrida, 1995a: 56; 'a general science of -+the archive, of everything that can happen to the economy of memory and to its substrates, traces, documents …': Derrida, 1995b, p. 34). These archival scholars, in their enthusiasm of this 'archival turn', are using the term 'archive' in poststructuralist sense, and are moving away from its traditional meaning(s), practices, and environments. They view 'the archive' as manifesting power, memory and evidence paradigms of past times and places (Stoler, 2002).

In the 1980s and 1990s, there was an intensive theoretical discourse about [1] the adoption of archival principles as 'respect des fonds' and 'provenance' (Bearman and Lytle, 1985; Carucci, 1992); and [2] the re-examinations of appraisal theory, instigated by Booms' (1987) evaluation of appraisal. The discourse revolved especially around how acquisition of archives by heritage institutions could represent society or social justice. Renewed attention to macro-appraisal theories (for the first time expressed by Hermann Meinert (1939)) and development of documentation strategies are expressions of that discourse. Acquisition was (and is) subject of theoretical (and practical!) scrutiny as it was challenged with rising amounts of information and a proliferation of information objects and records created by new technologies (Samuels, 1991, 1992; Duranti, 1994; Menne-Haritz, 1994; Brown, 1995; Cook, 2005). This discourse is still going strong: Shilton and Srinivasan (2007) and Huvila (2008), for instance, apply participatory design ideas to appraisal, and define the concept of participatory appraisal. Van Bussel (2012c) embedded appraisal within information relevancy theories that emphasize the change in information relevance over time.

3. The archival theoretical frameworks

3.1. The postmodern and structurationist meta-view of the Records Continuum

3.1.1. The Records Continuum theory and its long-term contribution

In the mid-1990s, Frank Upward (1996, 1997) defined his records continuum theory and model, with additional models following several years later (Upward, 2000). The continuum theory is influenced by Australian postcustodial practices, postmodernist thinking, and the social theory of structuration (McKemmish, 2001, p. 346-347; McKemmish, 2017, p. 137). Upward was especially triggered by Giddens'

view that societies are shaped by individuals and their structures (or traces of memory, as Giddens (1984, p. 378) calls them). Unlike the linear theory of information, thinking along the continuum emphasizes the continuous change in the context of information 'in spacetime' (Upward, 2000, p. 117-119). According to Xiaomi (2003), the theory is trying to integrate records and archives management, which is correct and its original intention as is clearly defined in Upward (1996). Upward (2000, p. 117) claims that his theory (and its postmodern and structurationist motivation) represents 'a fullly-fledged paradigm shift in which a worldview is being replaced', for it ends the 'life cycle worldview' that is based on 'the separation of space and time'. This claim that the theory is a paradigm shift has been supported (Thomassen, 1999; Cook, 1997, 2000a, 2001; McKemmish, 2001), but it has been correctly put into perspective by Luciana Duranti (2001) and Charles Jeurgens (2014). It is, at least, an exaggeration, for thinking in a semiotic spacetime continuum was introduced by the pragmatic philosopher Charles Peirce in the late nineteenth century (Morrissey, 2002; see also Upward (2017), without recognizing the contradiction with his earlier statement).[5]

---

[4] The *structuration theory* (or concept) of Anthony Giddens (especially in*: The constitution of society: Outline of the theory of structuration*, Cambridge, Polity Press, 1984) is developed as a social theory that tries to comprehend human social behaviour by studying the interfaces between actors (agencies) and structures. Giddens believes that actors operate within contexts of rules resulting from social structures. These structures do not have inherent stability outside the human action that constructed them. Agents modify social structures by acting outside their constraints. Giddens proposes three kinds of structure in a social system: [1] *signification*, a codification of meaning in language and discourse; [2] *legitimation*, normative perspectives implemented as societal norms and values; and [3] *domination*, the ways power is applied in the control of resources. Those structures are met by three kinds of interaction: [1] the communication of meaning; [2] morality or sanction; and [3] power relations. Structures and interactions 'communicate' with each other using a matching set of three modalities: [1] interpretive schemes; [2] norms; and [3] facilities. The object of the structuration theory are the conditions which govern the continuity and/or dissolution of structures and types of structures. In 1981 F. Gerald Ham ('Archival Strategies for the Postcustodial Era', *The American Archivist*, Vol. 44, No. 3, pp. 207-216) presented *postcustodialism*, a set of archival strategies that featured a decentralized computer environment that realized easy and centralized access to complex and decentralized archives. Ham did not argue that archivists should stop managing custodial holdings, but that they needed strategies to navigate the complex realities of the twentieth century. David Bearman ('An indefensible bastion: Archives as a repository in the electronic age', Technical report, *Archives and Museum Informatics*, Vol 13 (1991), pp. 14-24) went into extremes when arguing that archivists should avoid taking any custody at all of electronic records. In a networked world, 'it doesn't matter much where records or users are', as long as archivists have intellectual control. This provocative statement was endorsed in Australia (F. Upward and S. McKemmish, 'Somewhere Beyond Custody: literature review', *Archives and Manuscripts*, Vol. 22 (1994), No. 1, 136-149), but was abandoned several years later by the National Archives of Australia. In the establishment of a digital preservation project, it was argued than that digital records ideally should be tranferred to archival repositories for custody. In 2017, physical custody of archives is the stated preference of most archival programs as a result of the acceptance of 'trusted digital repositories'. A. Cunningham, 'Postcustodialism', L. Duranti and P.C. Franks, *Encyclopedia of Archival Science*, Lanham, Rowman & Littlefield, 2015, pp. 274-278.

[5] Apart from Peirce's pragmatic thinking of a semiotic spacetime continuum, the concept of a records continuum can be dated to the 1950s when Ian Maclean, the Australian national archivist, stated that archival science should be directed toward studying the characteristics of records, record keeping systems, and classification processes. He promoted a view of a management continuum for records. See: F. Upward, 'In Search of the Continuum: Ian Maclean's 'Australian Experience' Essays on Recordkeeping', S. McKemmish and M. Piggott, *The Records Continuum: Ian MacLean and Australian Archives: first fifty years*, Clayton (Vict.), Ancora Press, 1994, pp. 110-130. In 1985, Canadian archivist Jay Atherton made the word 'continuum' explicit for a way of integrated management of all interrelated stages of records, pointing out the information management weaknesses of the lifecycle model. This model is based on the premise that the 'life' of a record can be divided into two distinct, separate stages of responsibility: that of records management (with creation, classification, maintenance and use, and disposition of records) and that of archives management (with acquisition, description, preservation, and reference and use of archival records). J. Atherton, 'From life cycle to continuum: Some thoughts on the records management-archives relationship', *Archivaria*, Vol. 21 (1985), Winter, pp. 43-51. Upward's theory is the culmination of thinking about a records continuum, philosophically enriched by postmodernisms and Giddens' structuration theory.

The continuum theory and model are based on four dimensions: create, capture, organize, and pluralize, corresponding with four steps of time-space distanciation mentioned by Giddens (1984, p. 298) in an analytic example (!). The dimensions of the continuum describe how organizational archives (and the records captured within them) are disembedded from their original context(s) of use to become a part of a collective memory and carried through spacetime. Their context is represented by the axes of evidentiality, transactionality, record keeping, and identity (Upward, 2005). The theory is not about the archives themselves, it is about the information management activities that add new contexts to them such as capturing them into systems, or adding metadata. The status of archives is interpreted as part of a continuum of activity related to known and unknown contexts, to known or unknown social, cultural, political, and legal processes. According to the theory, it is this metaview, these contexts that are vital to interpret and (potentially) understand the role and value of archives in past, present, and future (McKemmish et al, 2010). A continuum approach highlights from the beginning that archives are both current and historical, representing one of the core concepts of structuration: the duality of structures. Archives and their records are viewed as fixed in content and structure, linked to mutable, ever-broadening layers of metadata to clarify their meaning and to enable their accessibility and usability over time (McKemmish, 2001).

Marshall (2000) states that the most important focus of the theory are the multiple purposes of archives (in multiple contexts) over time. Visualizations of the records continuum theory explain it (in essence) as a context theory, emphasizing the ever-broadening layers of contextual descriptions attached to records and archives. The aim of the theory is to provide a framework for conceptualizing archives in multiple contexts over space and time. Creating archives starts before they are created by implementing their requirements in policies, systems, organizations, processes, and laws. These requirements need to be integrated into social and business processes and purposes. The theory is heavily indebted to Australian postcustodial practices (see note 4), Terry Cook's (1992, 1997, 2005) ideas about macro-appraisal, and especially to David Bearman's (1993ab, 1994, 1996 (with Wendy Duff)) work on evidence, transactionality, and systems thinking. The influence of Bearman's extremely complex and inconsistent paper 'Record Keeping Systems' (Bearman, 1993a) is largely responsible for the mentioned axes of 'transactionality' and 'evidentiality'.

The theory's most important contribution is its accentuation of the importance of context and contextualizing for understanding the 'contextual narrative' of archives in spacetime. It has become common thinking in archival science that this 'contextual narrative' is an absolute necessity for revealing meaning, for accessibility, and for usability. But despite this long-lasting contribution and its very valuable insights into the context of records, which have greatly influenced my thinking about archives, from its formulation onwards, the theory *itself* has been on very shaky grounds.

3.1.2. Criticism: omissions, comprehensibility, and philosophical foundations

To counter *omissions*, some revisions of the theory have been suggested. Terry Cook (2000b) suggested (quite sensibly) to separate evidence and memory into their own

axes. He also suggests adding a new dimension (besides Create, Capture, Organize, and Pluralize) for archives of private origin. A fifth dimension is also (convincingly) proposed by Yvon Lemay and Anne Klein (2014), namely that of the use ('l'exploitation') of archives. But adding new dimensions to the theory is inconsistent with its structurationist nature. It would break the theoretical link to the four steps of time-space distanciation mentioned by Giddens (1984, p. 298). These steps are the sole reason for the four dimensions of the Records Continuum theory. New dimensions eliminate the possibility to directly link the records continuum to Giddens' structuration theory. Karabinos (2015) created 'the shadow continuum' to correct an omission in the theory concerning archives stuck between dimensions.

Michael Piggott (2012), an Australian supporter of the theory, made several remarks about the theory's problematic *comprehensibility* and its abstract nature. He states (2012, p. 180) that 'the core texts are not always easy to understand' and that it is very difficult 'to comprehend the intended meaning of continuum writing'. More problematic is his contestation that the continuum model is an abstraction that relies 'on the viewer to draw a correct inference' (Piggott, 2012, p. 183). That is confirmed by Karabinos (2015, p. 14) who states that it is the reader to make conclusions on what the model attempts to visualize because the model is 'confusing and vague'. One could characterize this as a postmodernist expression, but it is, of course, problematic, for a model that seemingly cannot convey its meaning in a straightforward way is very difficult to test (Piggott, 2012, p. 185).

The *philosophical foundations* of the theory are also heavily criticized. Verne Harris (2004, p. 215-216) condemns, in quite strong terms, Sue McKemmish's (2001, p. 347) claim for the model as 'post-modern philosophical … thinking' and to be 'universal' as 'the worst case of misidentification', as 'a co-opting – or colonising – move designed to have us believe that what is a wild tiger is only a domestic cat' and the fact that she 'ignores the fact that postmodernisms seek relentlessly to disturb every totalising conceptual container'. Harris is opposing the (theoretically untenable) totalizing worldview of the theory that ignores existing differences in information and records management. Andrew Lau (2013, p. 200-204) finds the structurationist theoretical foundations inadequate. Using Manuel DeLanda's (2006) neo-assemblage theory and its different view of society[6], he analyses the continuum theory and reveals, for instance, the mechanistic view of society and social complexity that allows for the reductionist approach the theory needs to identify stabilized entities that create archives. Such a view, however, is only one way

---

[6] Manuel DeLanda's neo-assemblage theory is an elaboration of the ontological framework developed by the postmodernists Gilles Deleuze and Felix Guattari in: *Capitalisme et schizophrénie 2: Mille plateaux*, Paris, Les Editions de Minuit, 1980. The theory offers a bottom-up framework for analyzing social complexity by accentuating exchangeability, indefiniteness, and multi-functionality. Deleuze and Guattari's assemblage theory is an approach that stresses that entities are not fixed, not predetermined, and not stable in their ontology or location. Assemblages are formed through coding, stratification, and territorialization processes. An assemblage, consisting out of imaginative articulations among heterogeneous elements, defines the relationships with the bodies in and around it, and demonstrates social complexity. See also: J.D. Slack, J. Macgregor Wise, *Culture and Technology. A primer*, New York, Peter Lang, 20142. Delanda's starting point is his argument that assemblage theory is a reaction to the theory of organic totalities. In his opinion, all 'parts' have some independence regarding the assembled 'whole' they help to constitute. Although a 'whole' will change following the addition or removal of an individual 'part', the components themselves do not need to change as a consequence of the new (dis)assembly. Assemblages, though dynamic, are part of historical processes. DeLanda defines a reinterpretation of the concepts of Deleuze and Guattari that provides a robust theoretical framework for analyzing assemblages. For an overview: M. DeLanda, *Assemblage Theory*, Edinburgh, Edinburgh University Press, 2016.

to view society and social complexity, and a reductionist approach may not be possible (or might be difficult) in other views of society. The claim of 'universality' is, thus, nonsense. Lau also attacks the form of objectivity and impartiality that is suggested in the theory: 'a privileged transcendent vantage point in which the archivist using the records continuum can 'see' all of society and speculate how records move from immediate contexts of creation through capture, organization, and the ultimate plural view of societal recordkeeping' (Lau 2013, p. 202). Such objectivity and impartiality of the archivist is impossible as understanding of reality is never complete. It is always constrained by the observer's perspective and knowledge and always carries blind spots, as Donna Haraway (1988) already stated. It ignores the effects of the archivist's context, his social preoccupations, moral codes, preconceptions, and choices made in the process of 'archivization'. This is remarkable in a theory that is (essentially) a context theory. As David Greetham (1999) concluded, archives do not tell us the truth. They propose a constructed, prejudiced, sometimes an idealized historical image. They are politicized bodies of information, pretending to be neutral. Joan Schwartz and Terry Cook (2002) stated that 'the archive' is always a reflection or a justification for the society that created it. Archivists are important in capturing, appraising, and maintaining archives and have a large responsibility for the reflection 'the Archive' provides.

3.1.3. Daily practice
The records continuum theory proves very challenging in the daily practice of managing business processes of organizations. Afshar and Ahmad (2015) propose a hybrid model for records management (a combination of continuum and linear models), because of the problems to implement the records continuum theory and model. Jeurgens (2014) concluded (albeit related to appraisal) that in daily (professional) practice 'in spite of all efforts and even the firm belief held by some professionals that archival thinking has undergone a paradigm shift, there is still no revolutionary progress in solving the many puzzles of records management and archiving'. The suggestion of Karabinos (2015, p. 150) that by removing the theory's 'universality' claim testing would not be necessary is, in my opinion, accepting defeat in bringing theory into practice. It is revealing that (although Sue McKemmish (2017, p. 143-144) states that the model is 'widely used as … an implementation model, as exemplified through its use in the development of standards, metadata entity-relationship models and schemas, and best practice guidelines for the design of recordkeeping systems and appraisal programs') practical examples of implementation of the theory within daily organizational practice are genuinely missing. The (needless) 'invention' of 'recordkeeping informatics' to 're-figure a discipline in crisis' (that is: a records management that does not use the records continuum theory and model) (Upward et al, 2013; Evans et al, 2014) only confirms that there is no convincing evidence that the theory has been used by organizations to manage their records and archives.

3.2. Digital diplomatics
3.2.1. The revitalization of traditional, proven methods in Digital Diplomatics
Not all archival scholars are following Foucault, Derrida or Giddens in their observations of 'the archive'. There is still a tradition, based on the 'old' diplomatic science, in which principles and concepts are 'universally valid', precisely defined, and 'objective' regardless of place. It provides a systematic method for the analysis of

the internal and external elements of documentary form, the circumstances of the writing, and the juridical nature of the fact that is communicated. It analyses the creation, form, and status of transmission of records, and the relationship with the facts represented in them and with their creator, in order to identify, evaluate, and communicate their 'true nature' (Duranti, 1998, p. 27). The primary focus of this tradition has been the 'record' (equated with the documents that were the subject of diplomatic science) and all the elements that it embodies. The content of the record is subject of its analysis, but also the relationships of the record and the persons, functions, procedures, acts, and the system that created them. The basic affirmations of this diplomatic tradition is that hypotheses and theories need to be empirically testable. Its philosophical roots lie in empiricism. Luciana Duranti is nowadays the most notable scholar within this tradition. She revitalized the traditional diplomatic methods and has argued for its relevance to electronic records and archives (Duranti, 1998, 2005, 2010a).

Digital diplomatics integrates traditional diplomatic techniques, concepts and methods with archival theory 'based on jurisprudence, the history and theory of administration, and an extensive and centuries old body of written reflection and experience' about the nature of records and archives within organizations (MacNeil, 2004, p. 205). Digital diplomatics emphasizes the importance of identifying evidence. For being used as evidence, records need to be authentic for only than it can be presumed their integrity has been maintained. To prove authenticity, the continuing identity and integrity of records and archives must be established. Identity and integrity of records allow to determine the who, what, where, when, and why, and establish 'perfection' in quality. Identity is revealed by documentary form or presentation. It is the whole of the distinguishing attributes that in combination uniquely characterize records. They have stable content and a fixed form, reveal together with the metadata layers of the organizational archive it belongs to, the legal, administrative, provenancial, procedural, technological, and documentary context, belong to identifiable organizations, persons or groups, are part of actions, are linked to related records, and are stored within the infrastructure of the organizational archive (Duranti and Jansen, 2011).

3.2.2. Computerized processing
Duranti has tried to broaden the types of records to which diplomatics could be effectively applied. In the InterPARES projects (1998-2018), she has applied diplomatic mechanisms to investigate the veracity of records in new, computerized environments (Ross, 2012). Within these projects, theory and methods are developed capable of ensuring the reliability, accuracy, and authenticity of electronic records and archives created in dynamic, experiential and interactive systems. InterPARES developed the Chain of Preservation, a series of continuous records-centric activities that contribute to the authenticity and preservation of records stored within the organizational archive. All activities that a record participates in are linked together. Any omission in a link deteriorates the ability of the chain to preserve the authenticity of records (and the archive they belong to), its ultimate objective (Jansen, 2015). According to Duranti (2009), it is possible that digital diplomatics may not be sufficient for dealing with challenges of increasingly complex digital environments, which might require that concepts, principles, and methods of other disciplines are evaluated by digital diplomatics. An expectation

I tend to agree with. As such, it is interesting to see how in digital records forensics the relationship between digital diplomatics and digital forensics is researched (Xie, 2011).

Digital diplomatics produces very detailed definitions and requirements for authentic electronic records that help business informatics in designing adequate ICTs for the organizational archive. That will be necessary: in an information deluge it is economically only viable to manage and store records in the organizational archive in a computerized way. The biggest contribution of digital diplomatics are the very detailed frameworks of authenticity and integrity requirements and its Chain of Preservation that allow (ultimately) for computerized processing and archiving of 'trusted' records. They have greatly influenced me in my ideas about the quality requirements for records and the information value chain. It is this contribution that makes digital diplomatics into a very interesting theoretical framework for EIM.

3.2.3. Theoretical problems

There are, however, some theoretical challenges with digital diplomatics. As Geoffrey Yeo (2017) points out, the equation of 'records' with the documents that were the subject of diplomatic science (particularly made in an English-speaking environment) may be a problem. The word 'record' was, until the late twentieth century, confined to countries whose legal and administrative systems are of English origin. It has no equivalent in other linguistic cultures. It was most certainly not common in the civil-law traditions based on Roman law in which diplomatic science has evolved (Yeo 2015). Yeo (2017) agrees that the equation seems to be correct for legal, textual records, for diplomatic science was primarily designed for their analysis. But is it correct for the new forms of record in the contemporary world that are largely non-legal and non-textual? Is it possible to apply diplomatic principles, techniques, and methods to these records, too? Joan Schwartz (1995, p. 54-55) does not think so and asserts that, at least for photographs, the 'extension of diplomatics from records of bureaucratic transactions created within the procedural rules, written or unwritten, of a juridical system to records of cultural actions and transactions' cannot be directly made and that 'the rigour of diplomatic criticism is undermined by the inherent ambiguity of the photograph'. Yeo (2017) also tends to answer those questions negatively and, although it is not explicitly addressed, he implies that the mentioned equation never has been researched adequately and that, for that reason, digital diplomatics is partly based on an unproven hypothesis. It is a serious allegation that cannot be refuted without researching the equation *itself*.

Even before digital diplomatics was born, there were issues with the reductive emphasis of diplomatic science on the relationship between record and juridical act. According to Leonard Boyle (1976) and Armando Petrucci (1995) that emphasis risked overlooking the contextual complexities of documents, their function and the power dynamics involved. Boyle (1976, p. 78) claims that the application of diplomatics demands intricate knowledge of the context of the object of analysis. A 'thorough competence in the language of the document' is needed, 'a knowledge of chronology', 'of local usages [and] conventions' and 'an easy familiarity with the methods, formulae, and practices … of the period and region'. Such knowledge is necessary to uncover the 'central reality' of the document to be analysed, but cannot

be gained by only using diplomatics. Petrucci (1995, p. 239) agrees with Boyle (1976) and states that 'the document is first and foremost evidence of a process entirely internal to its own making. And only the reconstruction of the process of documentation, of its articulations and reasons, can permit us to consider, with both greater insight and greater humility than before, the complex relationship between written documentation and the event that from time to time gave (or should have given) impulse to the process of documentation: what we call, in the language of diplomatics, the connection between action and documentation'.

Following these interpretations, Richard Brown (1997) challenged the strict (even narrow) contextual interpretation within digital diplomatics. He argued for a broader interpretation, in which social, cultural, ideological and other factors are considered, in addition to the very limited administrative-juridical context he observed in digital diplomatics at that time. And although digital diplomatists no longer perceive the juridical system as the only context for a record and recognize an extended range of contexts (legal, administrative, provenancial, procedural, technological, and documentary context (Duranti 2010b, p. 1596)), they have, as Yeo (2017) asserts, continued to emphasize the dominance of the legal context.

Digital diplomatics faces a contextual crisis. The context it captures is not enough in the long term to help users *understand* the wider social, cultural, and (inter-) organizational environment that generated the archive.

4. Concluding remarks

In this first part of the article, the problem EIM faces in contributing to organizational objectives and to defining business strategies was explored. To cope with the deluge of structured and unstructured information objects, EIM needs a theoretical foundation that effectively guides it in reaching business value. A possible solution for that problem is (following Smith and Steadman (1981)) 'the archive' and the records within it. Archival science has developed two overall theoretical frameworks relating to records and archives, but not focused on the (organizational or personal) construction of archives, the effects of (organizational) behaviour on their evolution, reaching organizational objectives, and designing business strategies. These two archival frameworks are based on philosophical traditions that are on opposite sides of the philosophical spectrum: postmodernism and empiricism. This shows itself especially in the very different forms of methodology used. The Records Continuum school uses deductive research methods, beginning with general a-priori concepts and, regardless of empirical data, deciding what to do with records and archives based on those concepts. Digital diplomatics uses inductive research methods, starting with empirical data and observations to find general principles about the subject.

The Records Continuum theory, characterized by structurationist and poststructuralist thinking, considers 'the archive' to be an epistemological and symbolic representation for the ways in which histories are constructed, organized, and narrated. An archive is a symbol of contestation, within which historical narratives, social power structures, and traditional meanings are challenged. This theory is not about records and archives themselves, but about their evolving

(especially societal) contexts within the dimensions of the theory that help to understand the meaning of their narrative in spacetime. It is used as a declarative model for the 'contextual narrative' of archives. Its philosophical foundation is weak, its comprehensibility problematic and its implementation in organizational practices debatable.

Digital Diplomatics, characterized by more empiricist, traditional, proven, and revitalized diplomatic techniques, methods, and concepts, is, in contrast, almost solely about the records and archives themselves and the relationships between them. It concentrates on the continuing identity and integrity of records to reveal documentary form, the attributes that characterize records and reveal its legal, administrative, provenancial, procedural, technological, and documentary functional context. It produces detailed definitions of authenticity and integrity that allow ICTs to recognize and realize 'trusted' records using its Chain of Preservation. The theoretical framework of Digital Diplomatics can be used as a declarative model for the (technical) generation of records, the relations between them, and their integrity and authenticity. It seems to be partly based on an unproven hypothesis and it faces a contextual crisis because the context it captures is not enough to understand the wider social, cultural, and (inter-) organizational environment that generated the archive.

The focus of both theories lays on the cultural (or historical) value (Records Continuum theory) and the evidential value (Digital Diplomatics) of archives. Both are important values that can be used to improve the way records and archives are managed and used within organizations. They offer interesting insights for EIM. The Records Continuum theory emphasizes the importance of context for a 'reconstruction of the past', extremely important for realizing accountability, governance, and compliance (Van Bussel, 2012b, 2016). Digital Diplomatics offers tools and frameworks to improve the authenticity and integrity of records to allow them to be used as evidence. Both theories offer, their theoretical weaknesses nothwithstanding, convincing arguments for the value of archives and records for organizations. But they have not succeeded in linking these values to the challenges of reaching organizational objectives, designing business strategies, and constructing archives in a way that offers EIM the possibility to do so effectively. Both theories do not explain how and why the archive is as it is.

To emphasize the organizational value of archives, there is, I think, enough space for another theoretical view: an *organizational* one, the view of the 'Archive-as-Is', a pragmatic view on archives and records, their genesis, construction, use, and continuous management in the everyday life of people and organizations. A view that can be used as a declarative model for understanding the archive 'as-it-is', how it has been designed, constructed, processed, manipulated, and managed as a valuable business resource within EIM. A view that explains how it has 'grown' to be the archive that the organization or the person that generated it, wants it to be, with all distortions consciously and unconsciously embedded within it.

Archives are constructed in organizational settings and are the result of organizational behaviour, business processes, and predetermined rules and regulations. Cultural and social preconceptions, deviant behaviour, and (conscious

or unconscious) negligence are influencing decision making within organizations and affect EIM in its management of records and archives. Researching the genesis of organizational (or personal) archives, the records within them, and their fundamental components is necessary to understand them, to contextualize them, and to use them for reaching organizational objectives, the design of business strategies, and the increase of business value. That is what they were primarily made for....

Literature

Afshar, M., & Ahmad, K. (2015). A new hybrid model for electronic records management. *Journal of Theoretical and Applied Information Technology, 81*(3), 489-495.

Alalwan, J.A., & Weistroffer, H.R. (2012). Enterprise content management research: a comprehensive review. *Journal of Enterprise Information Management, 25*(5), 441-461.

Baets, W. (1992). Aligning information systems with business strategy. *The Journal of Strategic Information Systems, 1*(4), 205-213.

Bearman, D. (1993a). Record Keeping Systems. *Archivaria, 36*, 16-36.

Bearman, D. (1993b). Functional requirements for record keeping systems. *Archives & Museum Informatics, 7*(2), 3-5.

Bearman, D. (1994). *Strategies for Managing Records in Contemporary Organizations*. Pittsburgh: Archives & Museum Informatics.

Bearman, D., & Duff, W. (1996). Grounding archival description in the functional requirements for evidence. *Archivaria, 41*, 275-303.

Bearman, D., & Lytle, R. (1985). The Power of the Principle of Provenance. *Archivaria, 21*(1), 14-27.

Benbasat, I., Goldstein, D.K., & Mead, M. (1987). The Case Research Strategy in Studies of Information Systems. *MIS Quarterly, 11*(3), 369-386.

Booms, H. (1987). Society and the formation of a documentary heritage. Issues in the Appraisal of Archival Sources. *Archivaria, 24*, 69-107.

Boyle, L. (1976). Diplomatics. In J.M. Powell (Ed.), *Medieval Studies. An Introduction* (pp. 69-101). Syracuse: Syracuse University Press.

Brothman, B. (1999). Declining Derrida. Integrity, tensegrity and the preservation of archives from deconstruction. *Archivaria, 48*, 64-88.

Brothman, B. (2001). The past that archives keep. Memory, history, and the preservation of archival records. *Archivaria, 51*, 41-80.

Brown, R. (1995). Macro-Appraisal Theory and the Context of the Public Records Creator. *Archivaria, 40*, 40-74.

Brown, R. (1997). Death of a Renaissance Record-Keeper: The Murder of Tomasso da Tortona in Ferrara 1385. *Archivaria, 44*, 1-43.

Burton, A. (ed.) (2005). *Archive stories. Facts, fictions and the writing of history*. Durham, NC: Duke University Press.

Campt, T. (2012). *Image matters. Archive, photography, and the African diaspora in Europe*. Durham: Duke University Press.

Carucci, P. (1992). Archival science today. Principles, methods and results. In O. Bucci (ed.), *Archival science on the threshold of the Year 2000: Proceedings of the international conference, Macerata, 3-8 September, 1990* (pp. 55-68). Ancona: University of Macerata.

Chaki, S. (2015). *Enterprise Information Management in Practice*. New York: Apress.

Chakrabarty, D. (2000). Subaltern studies and postcolonial historiography. *Nepantla: Views from the South 1*(1), 9-32.

Cook, T. (1992). Mind Over Matter: Towards a New Theory of Archival Appraisal. In B. L. Craig (ed.), *The Archival Imagination. Essays in Honour of Hugh A. Taylor* (pp. 38-70). Ottawa: ACA.

Cook, T. (1997). What is past is prologue. A history of archival ideas since 1898, and the future paradigm shift. *Archivaria, 43*, 17-63.

Cook, T. (2000a). Fashionable Nonsense or Professional Rebirth: Postmodernism and the Practice of Archives. *Archivaria, 51*, 14-35.

Cook, T. (2000b). Beyond the Screen: The Records Continuum and Archival Cultural Heritage. In L. Burrows (ed.), *Beyond the Screen: capturing corporate and social memory* (pp. 8-21). Melbourne: Australian Society of Archivists.

Cook, T. (2001). Archival science and postmodernism. New formulations for old concepts. *Archival Science, 1*(1), 3-24.

Cook, T. (2005). Macroappraisal in Theory and Practice: Origins, Characteristics, and Implementation in Canada, 1950-2000. *Archival Science, 5*(2-4), 101-161.

Cvetkovich, A. (2003). *An archive of feelings. Trauma, sexuality, and lesbian public culture.* Durham: Duke University Press.

Davenport, T.H., & Prusak, L. (1997). *Information ecology: Mastering the information and knowledge environment.* New York: Oxford University Press.

DeLanda, M. (2006). *A New Philosophy of Society: Assemblage Theory and Social Complexity.* New York: Continuum.

Derrida, J. (1995a). *Mal d'Archive: Une Impression Freudienne.* Paris: Galilée.

Derrida, J. (1995b). Archive Fever. A Freudian impression. *Diacritics, 25*(2), 9-63.

Dunbar, A.W. (2006). Introducing critical race theory to the archival discourse. Getting the conversation started. *Archival Science, 6*(1), 109-129.

Duranti, L. (1994). The concept of appraisal and archival theory. *The American Archivist, 57*(2), 328-344.

Duranti, L. (1998). *Diplomatics. New uses for an old science*, Lanham and London: The Scarecrow Press.

Duranti, L. (2001). The impact of digital technology on archival science. *Archival Science, 1*, 39-55.

Duranti, L. (2005). The long-term preservation of accurate and authentic digital data: the INTERPARES project. *Data Science Journal, 4*, 106-118.

Duranti, L. (2009). From Digital Diplomatics to Digital Records Forensics. *Archivaria, 68*, 39-66.

Duranti, L. (2010a). Concepts and principles for the management of electronic records, or records management theory is archival diplomatics. *Records Management Journal, 20*(1), 78-95.

Duranti, L. (2010b). Diplomatics. In M.J. Bates and M.N. Maack (eds.), *Encyclopedia of library and information sciences*, (3rd ed.). Boca Raton: CRC Press.

Duranti, L., & Jansen, A. (2011). Authenticity of Digital Records. An Archival Diplomatics framework for Digital Forensics. In W. Castelnova and E. Ferrari (eds.), *Proceedings of the 5th European Conference of Information Management and Evaluation* (pp. 134-139). Como: Academic Publishing Limited.

Evans, J., B. Reed, H. Linger, S. Goss, D. Holmes, J. Dobrik. B. Woodyat, & S. Henbest (2014). Winds of Change. A recordkeeping informatics approach to information management needs in data-driven research environments. *Records Management Journal, 24*(3), 205-223.

Faulkhead, S. (2009). Connecting through records: Narratives of Koorie Victoria. *Archives and Manuscripts, 37*(2), 60-88.

Ferro, N., & G. Silvello (2013). NESTOR: A formal model for digital archives. *Information Processing & Management, 49*(6), 1206-1240.

Foucault, M. (1975). *Surveiller et punir: Naissance de la prison*. Paris: Gallimard.

Foucault, M. (1976, 1984). *Histoire de la Sexualité*, Vol. I (1976), Vol. II-III (1984). Paris: Gallimard.

Foucault, M. (1992). *L'Archéologie du savoir*. Paris: Gallimard. (1st ed.: 1969)

Galin, J., & J. Latchaw (2001). Theorizing Raw Archive. A new paradigm for academic scholarship and publication'. In J.F. Barbar and D. Grigar (eds.), *New Worlds, New Words: Exploring Pathways for Writing about and in Electronic Environments* (pp. 279-306). Cresskill, NJ: Hampton Press.

Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. Cambridge (UK): Polity Press.

Gilliland, A.J., McKemmish, S., & Lau, A.J. (eds.) (2016). *Research in the Archival Multiverse*. Clayton, VIC: Monash University Publishing. http://dx.doi.org/10.26530/OAPEN_628143

Greetham, D. (1999). Who's in, who's out. The cultural politics of archival exclusion. *Studies in the Literary Imagination*, 32(1), 1-28.

Groth, P.T. (2007). *The origin of data. Enabling the determination of provenance in multi-institutional scientific systems through the documentation of processes*. Southampton: University of Southampton.

Haraway, D. (1988). Situated Knowledges: The Science question in feminism and the privilege of partial perspective. *Feminist Studies, 14*(3), 575-599.

Harris, V. (2004). Concerned with the writings of others: Archival canons, discourses and voices. *Journal of the Society of Archivists, 25*(2), 211-220.

Hausmann, V., S.P. Williams, C.A. Hardy, & P. Schubert (2014). Enterprise information management readiness: A survey of current issues, challenges and strategy. *Procedia Technology, 16,* 42-51.

Huvila, I. (2008). Participatory archive. Towards decentralised curation, radical user orientation, and broader contextualization of records management. *Archival Science, 8*(1), 15-36.

Huyssen, A. (2000). Present Pasts: Media, Politics, Amnesia. *Public Culture, 12*(1), 21-38.

Ismail, A., & A. Jamaludin (2009). Towards establishing a framework for managing trusted records in the electronic environment. *Records Management Journal, 19*(2), 135-146.

Jansen, A. (2015). Chain of Preservation. In L. Duranti, P.C. Franks (eds.), *Encyclopedia of Archival Science* (pp. 133-136). London: Rowman and Littlefield.

Jeurgens, C. (2014). *The target keeps moving. Appraisal in a continuum*. Paper presented at the conference 'Paradigm shift! Shifting minds and practices in record keeping?' (Nationaal Archief Netherlands, Den Haag, January 2014). Retrieved from https://openaccess.leidenuniv.nl/bitstream/handle/1887/32027/Jeurgens%20target%20keeps%20moving.def.pdf?sequence=1

Joyce, P. (1999). The Politics of the Liberal Archive. *History of the Human Sciences, 12*(2), 35-49.

Kaplan, E. (2000). We are what we collect, we collect what we are. Archives and the construction of identity. *The American Archivist, 63*, 126-151.

Karabinos, M.J. (2015). *The Shadow Continuum. Testing the Records Continuum Model through the Djogdja documents and the Migrated Archives*, Leiden (PhD).

Ketelaar, E. (1999). Archivalisation and archiving. *Archives and Manuscripts, 27*, 54-61.

Ketelaar, E. (2000a). Archivistics research saving the profession. *The American Archivist, 63*, 322-340.

Ketelaar, E. (2002). Archival temples, archival prisons. Modes of power and protection. *Archival Science, 2*(3-4), 221-238.

Ketelaar, E. (2017). Archival turns and returns. Studies of the Archive. In A.J. Gilliland, S. McKemmish, A.J. Lau (eds.), *Research in the Archival Multiverse* (pp. 228-268). Clayton, VIC: Monash University Publishing. http://dx.doi.org/10.26530/OAPEN_628143

Lau, A.J. (2013). Collecting Experiences. Doctoral Dissertation. Los Angeles: University of California. Retrieved from http://escholarship.org/uc/item/9f8572zp

Lemay, Y., & A. Klein (2014). Les archives définitives: un début de parcours. Revisiter le cycle de vie et le Records continuum. *Archivaria, 77*, 73-102.

Lynch, M. (1999). Archives in Formation. Privileged spaces, popular archives and paper trails. *History of the Human Sciences, 12*(2), 65-87.

Mackenzie, A. (1997). The mortality of the virtual real-time, archive and dead-time in information networks. Convergence: *The International Journal of Research into New Media Technologies, 3*(2), 59-71.

MacNeil, H. (2004). Contemporary Archival Diplomatics as a Method of Inquiry. Lessons learned from Two Research Projects. *Archival Science, 4*(3), 199-232.

Manoff, M. (2001). The symbolic meaning of libraries in a Digital Age. *Portal: Libraries and the Academy, 1*(4), 371-381.

Manoff, M. (2004). Theories of the Archive from Across the Disciplines. *Portal: Libraries and the Academy, 4*(1), 9-25.

Marshall, P. (2000). Life Cycle Versus Continuum - What Is the Difference? *Information Quarterly, 16*(2), 20-25.

Marcus, G.E. (1998). The Once and Future Ethnographic Archive. *History of the Human Sciences, 11*(4), 49-64.

McAfee, A. (2006). Enterprise 2.0: The dawn of emergent collaboration. *MIT Sloan Management Review, 47*(3), 21-28.

McDonald, T.J. (1996). Introduction. In T.J. McDonald (ed.), *The Historic Turn in the Human Sciences* (pp. 2-16). Ann Arbor: University of Michigan Press.

McKemmish, S. (2001). Placing records continuum theory and practice. *Archival Science, 1*(4), 333-359.

McKemmish, S.M., F.H. Upward, & B. Reed (2010). Records Management Continuum. In M.J. Bates and M.N. Maack (eds.), *Encyclopedia of Library and Information Sciences* (3rd ed.) (pp. 4447-4459). London: Taylor & Francis.

McKemmish, S. (2017). Recordkeeping in the Continuum. An Australian tradition. In A.J. Gilliland, S. McKemmish, A.J. Lau (eds.), *Research in the Archival Multiverse* (pp. 122-160). Clayton, VIC: Monash University Publishing. http://dx.doi.org/10.26530/OAPEN_628143

McKemmish, S., L. Iacovino, E. Ketelaar, M. Castan, & L. Russell (2011). Resetting relationships. Archives and Indigenous human rights in Australia. *Archives and Manuscripts, 39*(1), 107-114.

McLeod, J., & E. Lomas (2015). Records Management. In L. Duranti, P.C. Franks (eds.), *Encyclopedia of Archival Science* (pp. 346-350). London: Rowman and Littlefield.

Meinert, H. (1939). Die Aktenwertung. Versuch einer methodologischen Zusammenfassung. *Mitteilungsblatt der preußischen Archivverwaltung, 6*, 103-110.

Menne-Haritz, A. (1994). Appraisal or documentation. Can we appraise archives by selecting content? *The American Archivist, 57*, 528-542.

Morrissey, F. (2002). Introduction to a semiotic of scientific meaning, and its implications for access to scientific works on the Web. *Cataloguing & Classification Quarterly, 33*(3/4), 67-97.

Naidoo, K., & Nsibirwa, Z. (2015). The information behaviour of managers in the Msunduzi municipality's business units. *Innovation: Journal of Appropriate Librarianship and Information Work in Southern Africa, 51*, 3-21.

Nel, M.A. (2015). *Information needs, information seeking behaviour and information use behaviour of researchers at the Faculty of Veterinary Science, University of Pretoria, and how these needs are being met by the information support delivered by the Jotello F Soga Library*. Pretoria: University of Pretoria.

Nesmith, T. (1999). Still fuzzy, but more accurate. Some thoughts on the 'ghosts' of archival theory. *Archivaria, 47*, 136-150.

Nesmith, T. (2002). Seeing archives. Postmodernism and the changing intellectual place of archives. *The American Archivist, 65*, 24-41.

Okoli, C., & K. Schabram (2010). *A guide to conducting a systematic literature review of information systems research*. Sprouts: Working Papers on Information Systems. Retrieved from https://www.researchgate.net/publication/228276975_A_Guide_to_Conducting_a_ Systematic_Literature_Review_of_Information_Systems_Research

Orlikowski, W.J., & J.J. Baroudi (1991). Studying Information Technology in Organizations: Research Approaches and Assumptions. *Information Systems Research, 2*(1), 1-28.

Patnayakuni, R., & N. Patnayakuni (2014). Information security in value chains: A governance perspective. *Proceedings of the twentieth Americas conference on information systems (AMCIS 2014, Savannah), Association for information systems* (pp. 1-10). Retrieved from https://pdfs. semanticscholar.org/e4af/737191a4d6ecb90c990fd94bd4545e0f56cc.pdf.

Peppard, J., & J. Ward, J. (2016). *The Strategic Management of Information Systems: Building a Digital Strategy* (4th ed.). Hoboken: John Wiley & Sons.

Petrucci, A. (1995). The Illusion of Authentic History: Documentary Evidence. In A. Petrucci, *Writers and Readers in Medieval Italy: Studies in the History of Written Culture* (Chapter 10, pp. 236-250). New Haven: Yale University Press.

Piggott, M. (2012). Two cheers for the records continuum. In M. Piggott, *Archives and Societal Provenance: Australian Essays* (Chapter 12, pp. 175-195). Oxford: Chandos Publishing.

Rawson, K. (2009). Accessing transgender // desiring queer(er?) archival logics. *Archivaria, 68*, 123-140.

Redman, T.C. (2004). Data: un unfolding quality disaster. *DM Review, 8*, 22-23 and 57.

Ross, S. (2012). Digital Preservation, Archival Science and Methodological Foundations for Digital Libraries. *New Review of Information Networking, 17*(1), 43-68.

Samuels, H. (1991). Improving our disposition: Documentation Strategy. *Archivaria, 33*, 125-140.

Samuels, H. (1992). *Varsity Letters: Documenting Modern Colleges and Universities*. Lanham, MD: Scarecrow Press.

Schwartz, J.M. (1995). 'We Make Our Tools and Our Tools Make Us': Lessons from Photographs for the Practice, Politics, and Poetics of Diplomatics. *Archivaria, 40*(Fall), 40-74.

Schwartz, J.M., & T. Cook, (2002). Archives, records, and power: the making of modern memory. *Archival science, 2*(1-2), 1-19.

Serova, E. (2012). Enterprise Information Systems of new Generation. *The Electronic Journal of Information Systems Evaluation, 15*(1), 116-126.

Shetty, S., & E.J. Bellamy (2000). Postcolonialism's Archive Fever. *Diacritics, 30*(1), 25-48.

Shilton, K., & R. Srinivasan (2007). Participatory appraisal and arrangement for multicultural archival collections. *Archivaria, 63*, 87-101.

Smith, G.D., & L.G. Steadman (1981). Present value of corporate history. *Harvard Business Review, 6*(59), 164-173.

Srinivasan, M., & A. Dey (2014). Linking ERP and e-business to a framework of an integrated e-supply chain. In F.Z. Martínez-López (ed.), *Handbook of strategic e-business management* (pp. 281-305). Berlin and Heidelberg: Springer.

Stoler, A.L. (2002). Colonial Archives and the Arts of Governance. On the Content in the Form. In C. Hamilton, V. Harris, M. Pickover, G. Reid, R. Saleh, and J. Taylor, (eds.), *Refiguring the Archive* (pp. 83-100). London: Kluwer Academic Publishers.

Stoler, A.L. (2009). *Along the Archival GraIn Epistemic Anxieties and Colonial Common Sense*. Princeton: Princeton University Press.

Sundqvist, A., & P. Svärd (2016). Information culture and records management: a suitable match? Conceptualizations of information culture and their application on records management. *International Journal of Information Management, 36*(1), 9-15.

Thomassen, T.H.P.M. (1999). Paradigmatische veranderingen in de archiefwetenschap. In P.J. Horsman, F.C.J. Ketelaar, T.H.P.M. Thomassen (eds.), *Naar een nieuw paradigma in de archivistiek* (pp. 69-79). Den Haag: Stichting Archiefpublicaties.

Toffler, A. (1970). *Future shock*. New York: Random House.

Turner, V., J.F. Gantz, D. Reinsel, & S. Minton (2014). *The digital universe of opportunities: Rich data and the increasing value of the internet of things*. Framingham, Ma.: IDC.

Upward, F. (1996). Structuring the records continuum, part one. Postcustodial principles and properties. *Archives and Manuscripts, 24*(2), 268-285.

Upward, F. (1997). Structuring the records continuum, part two. Structuration theory and recordkeeping. *Archives and Manuscripts, 25*(1), 10-35.

Upward, F. (2000). Modelling the continuum as paradigm shift in recordkeeping and archiving processes, and beyond - personal reflection. *Records Management Journal, 10*(3), 115-139.

Upward, F. (2005). Records continuum. In S. McKemmish, M. Piggott, B. Reed and F. Upward (eds.), *Archives: recordkeeping in society* (pp. 197-222). Wagga Wagga: Charles Sturt University.

Upward, F. (2017). The archival multiverse and eddies in the spacetime continuum. In A.J. Gilliland, S. McKemmish, A.J. Lau (eds.), *Research in the Archival Multiverse* (pp. 198-227). Clayton, VIC: Monash University Publishing. http://dx.doi.org/10.26530/OAPEN_628143

Upward, F., B. Reed, G. Oliver, & J. Evans (2013). Recordkeeping informatics: re-figuring a discipline in crisis with a single minded approach. *Records Management Journal, 23*(1), 37-50.

Van Bussel, G.J. (2012a). Enterprise 2.0., Accountability and the necessity for digital archiving. In W.F. Riekert and I. Simon (eds.), *Information in e-Motion. Proceedings BOBCATSSS 2012 - 20th International Conference on Information Science. Amsterdam, 23-25 January 2012* (pp. 82-86). Bad Honnef: Bock+Herchen Verlag .

Van Bussel, G.J. (2012b). Reconstructing the Past for Organizational Accountability. *The Electronic Journal of Information Systems Evaluation, 15*(2), 127-137. Retrieved from www.ejise.com.

Van Bussel, G.J. (2012c). *Archiving should be just like an Apple, en acht andere (nuttige?) stellingen*. Amsterdam: Amsterdam University Press.

Van Bussel, G.J. (2016). An Accountability Challenge: Capturing Records and Their Context in Enterprise Information Systems. In P. Silva, A. Guerreiro and R. Quaresma (eds.), *Proceedings of the 10th European Conference of Information Systems Management. ECISM 2016, Evora, Portugal, 8-9 September 2016* (pp. 204-211). Reading: ACPI.

Van Bussel, G.J., & F.F.M. Ector (2009). *Op zoek naar de herinnering. Verantwoordingssystemen, content-intensieve organisaties en performance*. Helmond: Van Bussel Document Services.

Van de Pas, J., & G.J. van Bussel (2015a). Privacy lost - and found? The information value chain as a model to meet citizens' concerns. *Electronic Journal of Information Systems Evaluation, 18*(2), 199-209.

Van de Pas, J., & G.J. van Bussel (2015b). Embedding Privacy in ICT Architectures. The citizen as public stakeholder in architecture development. In B. van der Sloot (ed.), *Proceedings of the Amsterdam Privacy Conference (21-26 October 2015)* (14 pages, incl. References, only available on USB). Amsterdam: APPR. Retrieved from http://www.vbds.nl/wp-content/uploads/2015/10/Van-de-Pas_-Van-Bussel.pdf.

Van de Pas, J., G.J. van Bussel, M. Veenstra, & F. Jorna (2016). Digital Data and the City. An exploration of the building blocks of a Smart City Architecture. In D.P. Baker, W. Evans (eds.), *Digital Information Strategies. From Applications and Content to Libraries and People* (Chapter 13, pp. 185-198). Waltham, MA: Chandos Publishing.

Vom Brocke, J., A. Simons, & A. Cleven (2011). Towards a business process-oriented approach to enterprise content management: the ECM-blueprinting framework. *Information Systems and e-Business Management, 9*(4), 475-496.

Voss, P.J., & M.L. Werner (1999). Toward a Poetics of the Archive: Introduction. *Studies in the Literary Imagination, 32*(1), i-viii.

White, K.L. (2017). Race and culture. An ethnic studies approach to archival and recordkeeping research in the United States. In A.J. Gilliland, S. McKemmish, A.J. Lau (eds.), *Research in the Archival Multiverse* (pp. 352-381). Clayton, VIC: Monash University Publishing,

Withers, C.W.J. (2002). Constructing 'The Geographical Archive'. *Area, 34*(3), 303-311.

Wood, S., K. Carbone, M. Cifor, A. Gilliland, & R. Punzalan (2014). Mobilizing records: re-framing archival description to support human rights. *Archival Science, 14*, 397-419.

Wurl, J. (2005). Ethnicity as provenance. In search of values and principles for documenting the immigrant experience. *Archival Issues, 29*(1), 65-76.

Xiaomi, A. (2003). An integrated approach to records management. *The Information Management Journal*, July-August, 24-30.

Xie, S.L. (2011). Building Foundations for Digital Records Forensics: A Comparative Study of the Concept of Reproduction in Digital Records Management and Digital Forensics. *The American Archivist, 74*, 576-599.

Yeo, G. (2007). Concepts of record (1): evidence, information, and persistent representations. *The American Archivist, 70*(2), 315-343.

Yeo, G. (2015). Record(s). In L. Duranti and P.C. Franks (eds.), *Encyclopedia of Archival Science* (pp. 315-319). Lanham: Rowman & Littlefield.

Yeo, G. (2017). Information, records, and the philosophy of speech acts. In F. Smit, A. Glaudemans and R. Jonker (eds.), *Archives in Liquid Times*. Den Haag: Stichting Archief Publicaties.

Yin, R.K. (2003). *Case Study Research. Design and methods* (3rd ed.). Sage: Thousand Oaks.

Yusof, Z.M., & R.W. Chell (2002). Towards a theoretical construct for records management. *Records Management Journal, 12*(2), 55-64.

42

GEERT-JAN VAN BUSSEL

The Theoretical Framework for the 'Archive-As-Is'. An Organization Oriented View on Archives

Part II. An Exploration of the 'Archive-As-Is' Framework[*]

1. Introduction

In Part I of this article, I presented the first part of this exploration into the problems Enterprise Information Management (EIM) experiences in managing structured and unstructured information objects. It dealt with the possibility of using records and archives as applicable concepts to find a solution for that problem. It became clear that EIM lacks an applicable theoretical framework to use records and archives in its attemps to facilitate business processes in reaching organizational objectives and designing business strategies. To find a usable theoretical framework, the existing two archival theoretical frameworks were discussed. The conclusion of that discussion was that both theories, theoretical weaknesses nothwithstanding, offer convincing arguments for the value of archives and records for organizations. Another conclusion was that both theories have not succeeded in linking these values to the realization of organizational objectives, designing business strategies, and constructing archives in a way that allows EIM to facilitate organizations effectively in those endeavors.

In this part, I will extensively discuss the theoretical framework of the 'Archive-as-Is'. I developed the theory as a pragmatic view on archives and records, their genesis, construction, use, and continuous management. The 'Archive-as-Is' is a declarative model for understanding the archive of an organization (or organizational chain), how it has been designed, created, processed, manipulated, and managed as a valuable business resource. This framework explains how the archive has 'grown' to be the archive that the organization or the person that generated it, wants it to be (in short: the 'Archive-as-Is').

An overview of the conceptual background of the theoretical framework will follow this introduction. After that I will elaborate on the assumptions on which the theoretical framework is based, followed with a graphical model of the framework. The next part will be an in-depth discussion of all components of the framework. This part of the article will be concluded with several concluding remarks, remarks about further research, and an acknowledgement section.

2. Conceptual background of the theoretical framework

I have developed the theoretical framework of the 'Archive-as-Is' primarily as an *organizational* theory on archives. As such, the focus of the framework is on the organizations (and/or persons) that create, process, manage, and preserve information objects, records and archives in their business processes and activities. The background of the theoretical framework presented here is directly influenced by archival science, but also by concepts, theories, and ideas from organization and information sciences, such as:

1. The sensemaking theories of Karl Weick (1979, 1995) and Brenda Dervin (2003), that guide research about the way people make sense of information objects and the way organizations address either uncertain or ambiguous situations. For sensemaking, records and archives are of crucial importance, because of their contextual nature;

2. Relevance theories (Saracevic, 2007ab), which argue that what causes information to be used, stored, kept, and preserved is its relevance to the user or the organization that generates or collects that information. Relevance is extremely important when attributing value to records and should be part of appraisal processes;

3. The situation theory (Barwise and Perry, 1983; Devlin, 1994), an information theoretic mathematical ontology developed to support situation semantics. Situations support (or fail to support) items of information. The theory is applicable to the analysis of information flows and information architecture, cooperative action, and ICT-design (Israel and Perry, 1991; Devlin and Rosenberg, 2008). Situations can be associated with transactions in business processes and can be used to analyze records and the context(s) surrounding them;

4. Andrew Pettigrew's (1979, 1990) ideas of the relationship between context and organizational development, in which reconstructing past contexts, processes, and decisions to discover patterns, underlying mechanisms and triggers, is extremely important when formulating strategies, but also for accountability, governance and compliance;

5. The knowledge chain model of Clyde Holsapple (Holsapple and Singh, 2001), which offers a framework for knowledge translation within organizations to realize organizational objectives. It can be applied to records and archives because of its process-oriented nature;

6.   The activity theory as used by Bonnie Nardi (Nardi and O'Day, 1999; Kaptelinin and Nardi, 2012), which offers valuable ideas about behaviour and technology. Nardi (1996, p. 13) states that activity theory 'sees people and things as fundamentally different. People are not reduced to 'nodes' or 'agents' in a system; 'information processing' is not seen as something to be modelled in the same way for people and machines'. Nardi's theory has been important for my interpretation of EIM and organizational behaviour.

The philosophical tradition that underlies this new framework is pragmatism, in which 'truth' is traced by its 'respective practical consequences' (James, 1907, p. 45). Thought is not meant to describe or mirror reality (James, 1909, chapter 1). Theories should have practical application (James 1907, p. 216) and are instruments in problem solving, which is exactly the kind of logic useful in continuously changing organizations. The ethics of pragmatism is practical: ethical theory without practice is 'intolerably academic'. It should be judged by practical use (Dewey and Tufts, 1908, p. v). Patricia Shields (1998, p. 197) called pragmatism 'the philosophy of common sense'. Charles Peirce's general concept of 'continuum' has been extremely important for my understanding of information management, for 'every general concept is, in reference to its individuals, strictly a continuum' (Hartshorne and Weiss, 1933, p. IV, 642). Just as with the concepts of other pragmatist philosophers, Peirce's continuum is not bound by spacetime. Pragmatism is, by definition, an approach based on spacetime realities (as is recognized by Upward, 2017). Peirce's highly complex concept of 'continuum' would have been a sound philosophical foundation for the Records Continuum theory, but it was not recognized as such. Peirce's ideas about 'continuum' were revitalized in late twentieth century mathematics (Zalamea, 2003).

3. Assumptions

The framework of the 'Archive-as-Is' is based on several assumptions. These assumptions are:

1.   In the theoretical framework of the 'Archive-as-Is', the information management function is a continuum. It does not make a distinction between records management and archives management (commonly made in archival practices). The Information management function (and its expression: EIM) needs to guarantee content, context, and structure of records and archives over time, even if these records or archives cease to be used in business, even if there are different organizations/organizational units or persons responsible for (parts of) the information management function, even as (parts of) an archive are no longer retained and irreparably destroyed, and even if there are multiple legitimate successors of the organization or persons that created the archive, including archival repositories (archival institutions). This (pragmatic) continuum is not bound by spacetime.

2.   Records pass through a (non-linear) lifecycle. They are created and will, in the end, be irreparably destroyed ('die') or indefinitely preserved ('live') in the organizational archive, continuously managed in EIM processes and

contextualized by metadata that capture changing contexts in organizational, social and personal circumstances. Hence, the lifecycle of records takes place within a continuum of management and context.

3. Archives are neither complete, nor neutral or objective sources of 'truth (Lane and Hill 2010). Although they are 'process bound information' (Cook, 1997, p. 48; Thomassen, 1999, p. 76) and 'a sediment of organizational' (or personal!) 'actions' (PIVOT, 1994), they are *constructed* bodies, configured to retain all those records organizations or persons *choose* to retain, enriched with all the metadata that are *allowed* to be included in metadata schedules. Archives are primarily used to reconstruct the past (for, for instance, accountability) (Van Bussel, 2012b). They retain (at a minimum) all records that, according to legal obligations, have to be kept for specified periods of time. Archives embed all preoccupations, moral codes and preconceptions entrenched in procedures, business processes, legislation, and social environments. They are subjective constructs (Greetham, 1999). Not all records are captured in the organizational archive: employees may decide to delete them prematurely, because they do not find them relevant, do not want them to be known to anyone, do not want them to become part of accountability processes, or out of deviant behaviour. Archives change constantly: new records are added daily, metadata are added or changed, and records that have reached the end of their retention period are removed from the archive and irreparably destroyed. Only a (small) part of the archive is preserved indefinitely for its 'historical value'. That part of the archive can only deliver a distorted view of the reality in which the creating organization functioned (Kaplan, 2000).

4. In the *Manual for the Arrangement and Description of Archives* (1896, Muller et al, 2003, p. 19) in its Statement 2, it is declared that an archive 'is an organic whole', a 'living organism, which grows, takes shape, and undergoes changes in accordance with fixed rules. ... The rules which govern the composition, the arrangement and the formation of an archival collection, therefore, *cannot be fixed by the archivist in advance*; he can only study the organism and ascertain the rules under which it was formed' (italics by GJvB). Although this is true for archives that are no longer a 'living organism' (as is stated in a footnote), there may arise a problem for archives that are: organizational archives as digital, constructed bodies *need* to be configured *in advance*. This means that the business rules that govern composition, arrangement, formation, and (even) method of description are defined *before* the archive as a construct is created. They do not have a 'life'; they do not 'grow organically'. It is one of the reasons why archivists need to participate in the configuration phases of digital archives. But what does it mean for the statement of Muller, Feith and Fruin about the archive as an 'organic whole' when the business rules that define an archive *need* to be fixed in advance and do not grow organically? I do not have an answer now, but it needs careful consideration and research.

5.   It is possible that archival repositories will be 'without walls' (Cook, 2007, p. 429-430), but the opposite is also true. In this age of big data, organizational chains, inter-organizational data warehouses, cloud computing, authentic registrations, and computer mediated exchange, the archival repository may

be changing into a 'hub' for access to the original organizational and personal systems or web-environments that have managed the archive from the moment of its creation (a postcustodial view: Acland, 1991; Bearman, 1993a; Upward and McKemmish, 1994). Charles Dollar (1992) stated that as the integrity of archives and records would be best preserved in its original ICT environment, the costs of proprietary systems would be extremely high, and technology obsolescence would make preservation extremely complex, management of archives would become unsustainable for any archival repository. Duranti's (2007, p. 464-465) argument is that a physical place is an absolute necessity to maintain the integrity of archives. It is necessary that 'the archival institution establish an architecture in which the records of all creating bodies, once received, can be put into clearly defined and stable relationships, and in which their broader context can be identified and the associations among the records never broken' (a custodial view). Even adherents that agree with Duranti's argument about the absolute importance of guaranteeing the authenticity of records have disagreed with her conclusion that this only can be achieved by taken physical custody of the archive by an archival repository (for a discussion: Cunningham, 2015). Both statements are ideological and not substantiated with convincing practical evidence. In the theoretical framework of the 'Archive-as-Is', it is not important whether archives are preserved by the organizations that created them (or their successors) or transferred to an archival repository, although the practical consequences for EIM are far-reaching.

6. Archivists are part of the information management function of organizations. They help organizations in configuring policies, procedures, business processes, and ICTs to shape the organizational archive and to implement laws and regulations for compliance and accountability. They assist in developing metadata schedules that try to capture organizational and environmental contexts. They play a crucial role in reconstructing the past and appraising, selecting, contextualizing, and preserving records within the organizational archive. When they are working with an archival repository, they are acquiring and preserving archives, contextualizing and relating them, and realizing access. But they do *not* shape an objective narrative of past occurrences in preserving and contextualizing archives. They need to acknowledge their own subjectivity and the impossibility of creating complete and objective organizational or personal archives. They are part in deciding which archives will be indefinitely preserved and are accountable for gaps, inconsistencies, and distortions in (and between) them. Archivists are not neutral, independent, and objective custodians of organizational, cultural or historical knowledge.

7. My definition of a record (in Part I of this article) allows the inclusion of information objects that are traditionally not known as records and have not been part of organizational archives. There are information objects that, as Jenkinson (2003, p. 342) stated, have become a record because 'someone decided to stick it into a file rather than the bin'. They are set aside and preserved, maybe out of a notion of potential future value (as Schellenberg, 2003, p. 11-16, stated), maybe because of subjective perceptions of employees. If an organization wants to preserve an ebook because it is perceived as extremely valuable for the organization (although it is *not* evidence or cultural heritage), according to my definition it can be considered a record.

4. The theoretical framework

4.1. The framework's components

The framework of the 'Archive-as-Is' consists out of five components (A-E). The components A, B, and C are aggregations of several elements. These three *defining components* define the management of records and archives:

> A. *the four dimensions of information, (primarily) about records themselves*. This component is an aggregation of four elements Quality (1), Context (Situational) (2), Relevance (3), and Survival (4);
>
> B. *the two archival principles, about the archive as a whole*. This component is an aggregation of the elements Provenance (5) en Context (Environmental) (6); and
>
> C. *the five requirements for information access, about the accessibility of records and archives for users*. This component is an aggregation of five elements: Findability (7), Availability (8), Perceivability (9), Intelligibility (10), and Contextuality (11).

The fourth component is an *operational* one, the information value chain (D) that implements the first three components.

The fifth component is the *behavioural* component (E): organizational behaviour influences the information management function and the decisions that are made within EIM about the management of the information value chain.

4.2. The framework's model

The framework's model is presented in Figure 1.

Explanation of the model

The first three components of the framework (A, B, and C) are to be implemented by an organization into the information value chain (D) as mandatory requirements from global legal, accountability, and professional frameworks. The information value chain will manage records and create the organizational archive using its five primary and five secondary processes. The chain is configured to realize the three components A, B, and C, but is also embedded by organizational behaviour (E) that affects the management of records and the creation of archives. The information value chain manages the organizational archive as it is created and will continuously contextualize it when situational, organizational, and social environments change. An organizational archive and its records are accessible for all employees within an organization, of course dependent on security authorizations. When an archive is not mandatory transferred to an archival repository and stays within the organization itself, access from outside users could be arranged using an access hub, maybe (but not necessarily) realized by an archival repository.

The model can also be viewed from the perspective of an archival repository. When an archive is transferred to or acquired by an archival repository, the information value chain (D) of the repository will manage it. The chain is configured to know which archives are accepted, how they are to be processed, contextualized, preserved and continuously checked. The first three components

of the theoretical framework (A, B, and C) define the implementation of the information value chain of the archival repository. Organizational behaviour (E) influences the behaviour of the archivists and their choices (in acquisition, contextualizing, preserving, etc.) are based on social, moral, and professional norms, codes and preconceptions. Archivists are continuously contextualizing the archive. The five requirements of information access (C) are very important for archival repositories. Repositories need to facilitate their users in realizing all requirements of information access and this means, in the end, implementing technologies to facilitate human-computer interaction.
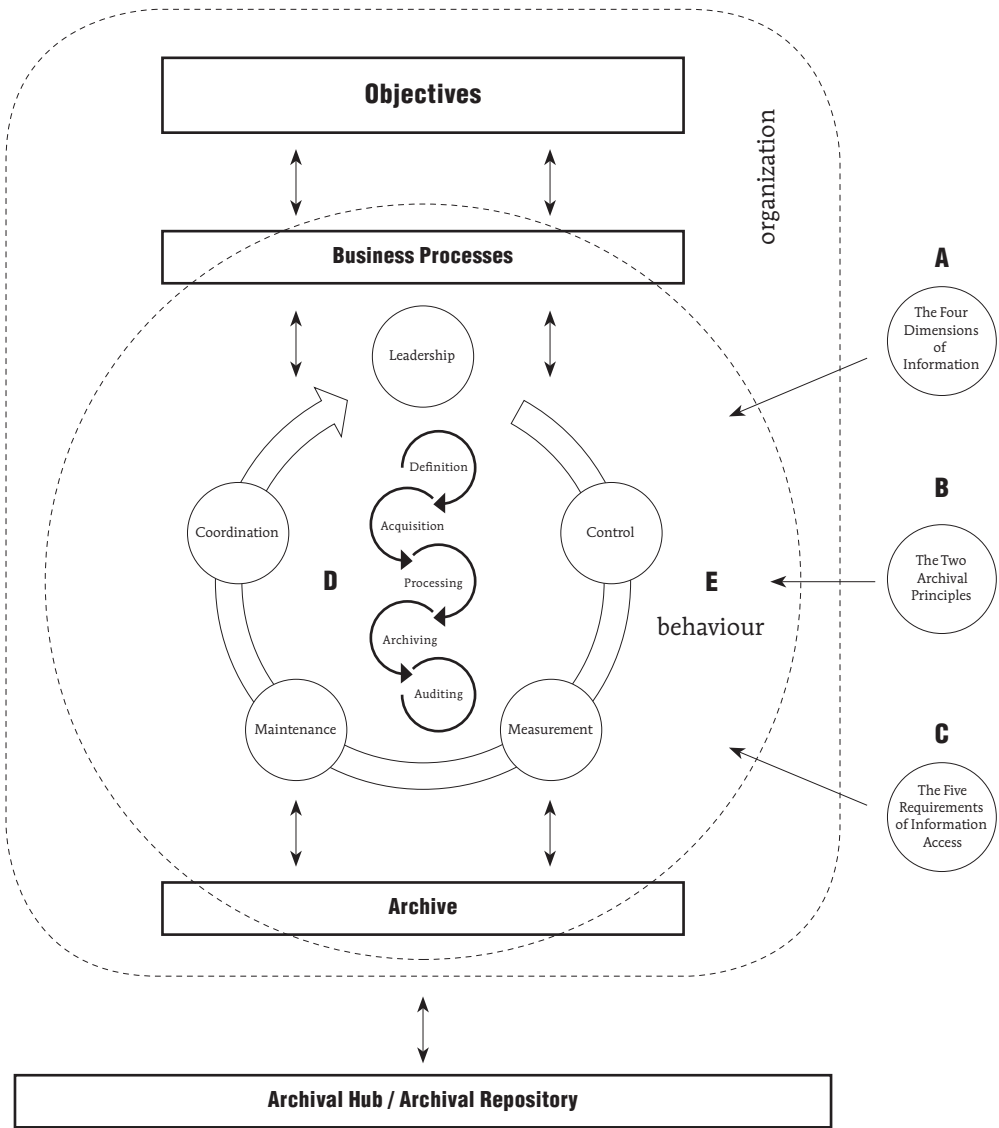
Figure 1. The Theoretical Framework of the 'Archive-as-Is'

4.3. The three defining components of the 'Archive-as-Is' (A, B, and C)

These components define the dimensions, principles, and requirements that must be met by organizations to realize EIM to be effective and to retain 'trusted' records that can be used to reconstruct the past. The defining components must be implemented as obligatory requirements in the lifecycle of records and the continuum of the information management processes facilitated by EIM. These three components are requirements for organizational records, archives, and their management, imposed on organizations by global legal, accountability, and professional frameworks.

4.3.1. The four dimensions of informatIon (A)

In complex computerized environments, the trustworthiness of records is constantly challenged. That is a problem, because records are meant to be (and are used as) evidence for organizational (or personal) policies, decisions, products, actions and transactions. Citizens, governments, and courts are making increasing demands for their trustworthiness (El Kharbili et al, 2008). Four dimensions of information allow for a reliable reconstruction of these policies, decisions, products, actions and transactions: quality, context, relevance, and survival (Van Bussel 2012a). These four dimensions are the four elements of the first component of the framework.

The *first* dimension, *Quality* (1), is about the quality requirements of records (according to assumption 7) and the 'information value chain', which will be discussed later as the fourth component of the framework. Van Bussel and Ector (2009, p. 181-214) describe this dimension based on an analysis of organization and information sciences literature about the quality requirements of data and information as well as the results of digital diplomatics research. Four quality requirements are recognized for records: *integrity* (they cannot be manipulated), *authenticity* (they present the required (and original) content and structure), *controllability* (they can be tested on integrity and authenticity) and *historicity* (their content, context and structure can be reconstructed at any moment in time). These four requirements realize the fixity of records. This means that they are (or can be reconstructed as) 'immutable mobiles' (Latour, 1990). Fixity is a necessity because records are meant for *later* consultation and are used repeatedly for the reconstruction of *past* happenings. Fixity enables users to trust records and to use them, for instance as evidence (Levy, 2001, ch. 2). The 'information value chain' ensures that records meet these quality requirements in spite of all necessary handling and guarantees that the necessary context is added, needed to allow for later sensemaking and to identify specific records. The requirements for this value chain are identical to those for organizational business processes, namely reliable time of delivery, effectiveness, efficiency, product quality, alignment of needs, product management, and compliance (Van Bussel and Ector 2009: 205).

The *second* dimension is *(Situational) Context* (2). According to Brenda Dervin (1997), context is an 'unruly beast'. 'There is no term that is more often used, less often defined, and when defined, defined so variously as context' (Dervin, 1997, p. 13-14). The concept has been attributed with many definitions, interpretations, and frameworks that can be divided into four classifications. There are interpretations that use context for defining and operating robotic activities in the

((very) near) future, like oracles (Devlin, 1991), for creating a situational environment for a user when using information, for adapting sofware applications to the personal context of the user, and for sensemaking of (the information in) social situations. This last interpretation of context is the subject of this second dimension of information. The context dimension of records is about the social situation (actions or transactions, cases, process flows, etc.) that generates them. This dimension captures the *situational context of individual records*. This situational context has some characteristics that are agreed upon in literature:

1. it is (in a phenomenological view) a complex social reality that (in a positivist view) will be captured as a simplified metadata construct that is a mere representation of that social reality (Penco, 1999);
2. it encapsulates records and situations to allow for sensemaking (Weick, 1979, 1995; Dervin, 2003);
3. it needs accurate documentation and definition (Groth, 2007);
4. it is in the past (Van Bussel, 2012b); and
5. it is necessary for the tracking and reconstruction of social situations, like business processes, policies, decisions, products, actions, and transactions (Groth, 2007; Self et al, 2007; Van Bussel and Ector, 2009).

The context of social situations provides meaning for the records generated within that situation (Weick, 1979, 1995; Dervin, 2003; Duranti, 1997a). To extract meaning out of situations (cases, process flows, decisions, etc.), EIM users need to gather knowledge of the individual organizational policies, decisions, products, actions or transactions for which records were generated (and their relationships) (Barwise and Perry, 1983; Devlin, 1994). The dimension of context captures data of the existing regulation(s) for the business process the records are part of, the business process itself, the structure of the specific case, the procedures by which records are generated, processed, and used, and their place in the information structure they belong to (Van Bussel and Ector, 2009, p. 215-260). This situational context of records is captured in metadata that try to generate an image of the specific action or transaction records are part of, the changes therein over time, their processing and use, and its management. These metadata have an unbreakable link with the records they belong to (Van Bussel, 2016).

The *third* dimension, *Relevance* (3), is an important concept in human communication and information management. As Saracevic (2007ab) explained, records are only relevant for users if they fit the context in which they are used, managed and retrieved. They need to be relevant for organizational or personal objectives of performance and accountability. They need to have pragmatic quality (Van Bussel, 2012a). A special kind of relevance is *appraisal*, determining the 'value', relevance, of records over time (Van Bussel and Ector 2009, p. 301-309). Appraisal is the complex (and quite subjective) evaluation of records to determine their economic, organizational, financial, fiscal, juridical, legal, societal, and historical relevance and to develop organizational or personal retention schedules. Such schedules define the periods of time that records should be kept or 'retained' (as, for instance, stated in law and regulations), including indefinite retention for records of 'enduring value' and the (not always mandatory) acquisition of organizational archives by archival repositories (Cox and Samuels, 1988). Appraisal is based on the

assumption that when a retention period has expired, records have lost their organizational, legal, and historical relevance and should be irreparably destroyed (Van Bussel, 2012a). For organizations of local, regional and national governments the subsequent selection and disposal of records are most often mandatory. Although not mandatory for non-governmental organizations, disposing of irrelevant records saves (potentially high) costs for retention and accessibility. Besides that, irrelevant records make organizations vulnerable to legal proceedings, for instance in the context of privacy law, fraud or corruption (Van Bussel and Henseler, 2013). The much disputed 'right to be forgotten' is an essential part of the discussion on the relevance of records (Mayer-Schönberger, 2009; Stupariu, 2015).

The *fourth* dimension of information concerns the *Survival* (4) of records over time. It pertains to the security and durability challenges, which have to be overcome to realize access, retrieval, and preservation of records in spacetime (Bearman, 2006). It stresses the importance of a reliable and durable ICT infrastructure to enable the continuous and secure storage of records. The features of this infrastructure are fragile and continuously influenced by the restructuring of organizations (Boudrez et al, 2005). The challenge of preservation is almost overwhelming. First, hard- and software configurations are always needed for accessing, retrieving and viewing information, which means that a solution for technological obsolescence should be available. Secondly, the large influx of information requires automated archiving and retrieval functionalities. The ICT infrastructure needs to adapt, transform, renew and grow, but this enhances the risks for obsolescence. Thirdly, records are of a diverse nature. There is a diversity of object types, operating systems and applications. The handling of this diversity is not self-evident, while it is, at the same time, not impossible to change the content of records, which endangers the trust in their reliability. Fourthly, records can only be reliably used, when they can be interpreted by users in their original situational context. A case-based review of this dimension has been offered by, among others, Hockx-Ju (2006).

4.3.2. The two archival principles (B)
I recognize two fundamental archival principles, an 'old' and a 'new' one, the principle of *Provenance* (5) and the principle of *(Environmental) Context* (6) respectively. Both principles are closely interrelated. It may even be difficult to differentiate between them as a result of the intermingling of both principles within archival scholarly literature. The principles are about the archive as a whole and, indirectly, about the records within it.

The 'old' archival principle of *Provenance* (5) is seen as the 'foundation of archival theory and practice' (Horsman, 1994, p. 51). This 'ambiguous concept' (Sweeney, 2008) has been a topic for scientific discourse since its introduction in the eighteenth and nineteenth centuries. It still is. According to Shelley Sweeney (2008, p. 194) 'over the years the principle has been introduced, reintroduced, applied in part, applied in full, studied, and debated without end'. Giovanni Michetti (2016) defines provenance (based on ICA definitions) as the relationship between archives and the organizations or persons 'that created, accumulated and/or maintained and used [them] in the conduct of personal or corporate activity'. It is also the relationship between them and the functions that generated their need. The word

'provenance' refers, hence, to 'the origins of an information-bearing entity or artifact' (Sweeney, 2008, p. 193). That is important, because archives 'should be arranged according to their provenance in order to preserve [its] context, hence, [its] meaning' (Michetti, 2016, p. 59). From its early history, the principle of provenance was meant, first, not to intermingle archives from different origins ('respect des fonds') and, second, to maintain the internal structure of an archive in its 'original order' ('archival bond') because it is a reflection of the functions of an organization or an individual. Both are needed for an archive to have evidential and informational value (Schellenberg, 2003; Posner, 1967; Horsman et al, 1998; Reilly, 2005).

Provenance has become a research object in other disciplines to see how it can be used and represented in different contexts. In computer science, the interpretation of provenance is that of data lineage, a description in the ownership history of how a data object was derived (Buneman et al, 2001). Records can become an aggregate of several information objects, may be stored in several locations, may be (part of) databases, documents, spreadsheets, or emails, may cross organizational borders, and may become part of one or more archives. Along the way, their origin and its logistic history may become obscure, may contain gaps, or may be lost (Puri et al, 2012). Systems are developed that trace and analyse provenance across distributed, networked environments, like Chimera in physics and astronomy, myGrid in biology and CMCS in chemical sciences (Simmhan et al, 2005). In visual analytics, it is recognized that the need to trace provenance extends beyond computing and into the realm of human analysis (Lemieux, 2016). In computer science, the focus is on individual items, while in archival science it usually applies to an archive or an aggregation of records. Tom Nesmith (1999) associates provenance with the social and technical processes of inscription, transmission, contextualization, and interpretation of archives, which account for their existence, characteristics, and continuing history. It broadens 'the idea of provenance ... to include its societal dimensions' (Nesmith, 2015, p. 286). It is a postmodernist interpretation that unmistakable intermingles provenance and context. Using the principle of provenance proves to be complex when there is a 'parallel provenance, 'two or more entities residing in a different context as establishing the provenance of [archives], even when they are involved in different kinds of action, for example creation and control' (Ketelaar, 2007, p. 186-187, based on Hurley (2005)).

The *object* of the principle of provenance is the (business process) archive of an organization or an organizational chain *as a whole* and the *structure of relationships* within that archive. It is *not* meant to contextualize archives. It *only* wants to ascertain that: [1] archives (or aggregations of records) can be traced back to their creator(s) and their creation, and [2] the 'archival bond' in which their records are embedded can be reconstructed (Duranti, 1997b). For EIM the principle means that metadata about the creation and logistic history of organizational archives are to be preserved and that their internal structure(s) must always be reconstructable. Nevertheless, tracing the history of individual records to safeguard the four dimensions of information seems to be necessary in digital environments (Cui and Widom, 2003). In reconsidering the archival principle of provenance, this is an important reason to add data lineage to the implementation of the principle.

(Environmental) Context (6), the second archival principle, is a 'new' principle. It is comparable to the 'ambience function' introduced by Chris Hurley (1995). Its *object* is *not* the archive, but the environmental circumstances that give the archive meaning and that allow for its interpretation. It defines and captures the surrounding influences of the archive in metadata. It is an 'outside' phenomenon 'even if it conditions meaning and, in time, its interpretation' (Duranti, 1997b, p. 217). This context captures metadata about the organizational, personal, and social environments of the archive, the environment the organization directly experiences and that modifies its responses (Pfeffer and Salancik, 1978, p. 72-74). It also concerns the organizational structure, the business process hierarchy, and the legal and regulatory environment in which the archive is generated. Eric Ketelaar (2000b) adds social-cultural influences from the wider organizational environment to that mix. His views are closely related to the sensemaking theories of Karl Weick (1979, 1995) and Brenda Dervin (2003). To capture a representation of these influences in metadata is, however, extremely complex.

No one disputes the contextuality of archives. But the boundaries of the principe of provenance have been stretched to include environmental context, neglecting the fact that the object of provenance is the archive, its internal structure of relationships, and its lineage. Its object is *not* the environment of the archive that allows for sensemaking. Michetti (2016, p. 59), thus, is incorrect in stating that the arrangement or archives according to their provenance preserves their 'context, hence [their] meaning'. It preserves their source, internal structure, and lineage, but not their context. The building blocks for the understanding and interpretation of archives are their environmental influences, their environmental context, in a very simplified way captured within archival metadata (Van Bussel, 2016).

Context is an axiom. But it has never been considered a principle within archival science, although an archive (and the records within it) without a context is a meaningless aggregation of data that cannot realize the organizational or cultural objectives archives are constructed or used for. I am applying the context principle(s) of Frege (1980 (1884)) and Wittgenstein (1961 (1922)) to archives and define the rule that an archive (and the records within them) can only have meaning within their environmental, surrounding influences. The principle of context expresses, thus, the rule, in short, to never ask for the meaning of an archive (or its records) in isolation, but only in its context. That context is what EIM needs to capture in metadata to ensure that archives can contribute to the realization of organizational objectives (Van Bussel, 2016).

The context *dimension* of a record is guided by the context *principle* of the archive in supplementing the situational context of a record with the environmental context of the archive. Both contexts help in reconstructing the situations that generate(d) records and the organizational, personal, cultural, economic, and/or social circumstances that determine(d) creation, management, and preservation of archives. Situations and surrounding archival influences are captured in a simplified way in metadata.

4.3.3. The five requirements for information access (C)

Almost twenty-five years ago, Michael Buckland (1991, p. 77) stated that 'access emerges as a recurrent theme' within information science, but information access is hardly conceptualized. In archival science, there is work done about the access to archives. It concentrates on access permissions, freedom of information, legal restrictions, and the arrangement of archives (Kozak, 2015; Thomassen et al, 2001). There are no overall concepts of information access in archival science. In information science, however, two theories modelling the concept of information access have been developed. Both theories have contributed to the understanding of its dimensions. None of these theories have explained what the facets, or require-ments of access are (McCreadie and Rice, 1999; Burnett et al, 2008). Kay Mathiesen (2014) recognized five facets of access, largely corresponding to the five requirements of information access I have defined.

Information access for users has to be realized *regardless* of technology, language, disability, or personal capabilities. Its importance is growing in an age of an expanding digital universe, expanding legal frameworks and organizational accountability, and changing notions of privacy, economy, literacy, and daily life. Because of its complexity, it can 'be a burden' (Mason, 1986, p. 10-11). I recognize five requirements for information access that *together* define if (potential) users have access to archives and records.

This *first requirement* is *findability* (7). It concerns the possibility an individual has to discover *where* records are created, published, kept, stored, or preserved. Finding something refers to locating something in a known space. So, finding records is not a search problem (which attempts to locate something in unknown spaces), but an EIM problem (Baker, 2013). Findability is an essential part of both social and organizational information architectures. These architectures try to ensure that users can find records easily in spaces where complexity, information overload, and unfamiliarity hamper findability (Resmini and Rosati, 2007). Such architecture is necessary because the inter-subjectivity between the person or organization that created and/or organized archives and records and the persons looking for the content of those archives and records complicates finding them (Berlin et al, 1993; Narayan and Olsson, 2013). Information architectures try to realize cognitive and informational continuity between different environments. That way, users do not have to shift constantly between different, often colliding patterns of information structuring (Resmini and Rosati, 2007). Finding-aids are of the utmost importance for users to find the archives and records they need.

The *second requirement* is *availability* (8). Even if archives and records are 'findable' (the potential user knows *where* they can be found), that does not mean they can be retrieved and be made 'available' at a certain moment in time. There may be barriers that could make obtaining records difficult or, even, impossible. There may be legal ownership restrictions that do not allow their availability. Archives may be deemed confidential by the organization that preserves it. Records may have been irreparably destroyed or may have disappeared. They may be in a repository that is hosted behind a pay wall. The ICTs needed to obtain them may not be available. Even if ICTs are available, it is not unlikely, especially when trying to retrieve 'older' records, that software cannot decipher the data formats originally used. Archives and records may

be deemed as not of enduring importance and, as such, not acquired by archival repositories or kept by their creating organizations. So, although a user knows where archives and records are ('they are findable'), he or she cannot obtain them ('they are not available').

When archives and records are findable *and* available, they should be *perceivable* (9), the *third requirement* of information access. It should be possible to perceive them, to hear, feel, smell, taste, or view their content. If potential users are disabled in ways that prohibit hearing, feeling, smelling, tasting, or viewing, there should be assistive and interactive technologies in operation that allow them to perceive records (Hill 2013). When records are heard, felt, smelled, tasted, and/or viewed, users have the *possibility* to gather their meaning (Jones 2011). It is only *possible*, for even if records are findable, available, and perceivable, that does not mean they are 'intelligible'. To ensure accessibility and usability at both perceptual and cognitive levels of human-computer interaction, designers of archival systems need to be constantly aware of such design issues and should integrate those issues in evaluating their designs (Kato and Hori, 2006).

The *fourth requirement* of information access is *intelligibility* (10). Perceivable records can be read, heard, felt, smelled, and/or viewed, without the user having the capabilities to *understand* them. Understanding is only possible if the information literacy capabilities of users enable them to do so. According to the Karlsruhe concept of comprehensibility, the most ideal level of intelligibility depends on six dimensions: simplicity, structure, correctness, motivation, concision, and perceptibility. If an information user cannot (completely) gather one (or more) of these dimensions, it becomes more difficult to understand the records (Göpferich, 2006). Facilitating intelligibility may be a burden for organizations (archival repositories among them), because even in very literate countries large minorities of the population can only read simple texts in their own language (OECD 2015). Those minorities may be less educated people, immigrants, untrained readers, or people with dyslexia, aphasia, intellectual or cognitive disabilities, learning disabilities, or neuropsychiatric disabilities. Much above the level of 'simple text' is for most of those people *unintelligible*. For that reason, for large minorities of the population accessing records will be problematic. To have access to ICTs will not solve the problem, which makes the dissemination of knowledge quite difficult.

The last, *fifth requirement*, is *contextuality* (11). Archives and their records may be findable, available, perceivable, and intelligible, but if their contextuality is in jeopardy, it may be impossible to reconstruct the situational and environmental context in which they were generated, used, and managed. This requirement is connected with the dimension of (situational) context (2) and the principle of (environmental) context (6) as it allows users to access archives and records in context. Archives and records have a specific meaning in the context in which they are (were) generated and used. If their situational and environmental context cannot be reconstructed by a user, the meaning they were meant to have at the moment of their creation or as a consequence of their use, will be lost. At that moment, they lose their function as reference, as evidence of actions and transactions, or as source of organizational knowledge. If that context is unavailable

or impossible to reconstruct, archives and records may be interesting for users, but only in their own context of information seeking (Kulthau, 2006). This requirement allows users to interpret the meaning of archives and records in a way that was intended by the organization or person that constructed the archive. That interpretation will not be complete and will be restricted by the metadata that were allowed to be captured during creation, use, management, and preservation of the archive and the records within it. What is done with that context by users is dependent on their (research) questions. They may try to find other contexts unconsciously embedded into the records or the archive, like Emmanuel Le Roy Ladurie (1975) did for Montaillou or Catarina Bruschi (2009) for the Waldensian heretics in the Languedoc.

The requirements of information access are defined from the viewpoint of the *users* of the archive and its records. For them to be useful for the user, they should be accessible. Meeting information access is one of the biggest challenges for EIM. The five requirements of information access define this challenge. It means that EIM will have to meet every requirement of information access, including all technologies needed for users to perceive records, including generation or maintenance of information architectures that allow users to quickly access archives, and including all contextual metadata for archives and records to allow for a reconstruction of the past.

4.4. The operational component of the 'Archive-as-Is': The information value chain (D)

The three defining components of the theoretical framework of the Archive-as-Is are to be implemented by organizations as mandatory requirements in the *operational component* of the framework: the *information value chain*. This chain of information processes, organized by EIM, realizes these components in the business processes of organizations. That way EIM assists these business processes to reach organizational objectives. EIM organizes the information value chain to identify, control, and manage archives, records, and ICTs in and between organizations. The chain ensures that the informational and evidential value of records is utilized in and between business processes to improve performance, privacy and security by safeguarding the four dimensions of information, the two archival principles, and the five requirements of information access (Van Bussel and Ector, 2009; Van Bussel, 2012ab). It is recognized that managing records is a critical source for competitive advantage (Holsapple and Singh, 2001). Michael Porter and Victor Miller (1985) point out that between organizations, differences in the management of information (thus, archives and records) have an effect on activities and lead to differences in their competitiveness.

The information value chain identifies ten distinct, generic processes and nineteen activities that an organization (an organizational chain and/or even a person) performs when managing its records. The chain is comprised of five primary processes, used to manipulate the organizational archive and its records, and five secondary processes that guide performance of the primary processes and their activities. These primary processes and their corresponding activities do not need to be performed in a strict pattern, but there can be various sequences and overlaps among them. The secondary processes influence these variations. In structuring the

Information Definition		Defining the four dimensions of information, the two archival principles and the five requirements of information access within organizational policies, procedures, rules, and systems.
Activity 1	*Configure*	Configuring policies, procedures, rules, and systems to implement the four dimensions of information, the two archival principles, and the five requirements of information access, using requirements of all activities of the information value chain.
Information Acquisition		Generating and/or acquiring records (and/or archives) from internal and external sources to make it suitable for subsequent use within specifically set procedures and conditions.
Activity 2	*Generate/receive*	Creating and receiving records (and/or archives).
Activity 3	*Identify*	Identifying records (and/or archives) and adding context.
Activity 4	*Capture*	Capturing records (and/or archives) in defined and configured information and archiving systems
Activity 5	*Store*	Store records (and/or archives) in information and archiving systems and making them suitable for subsequent use
Information Processing		Processing and analysing records (and/or archives) in business processes to get work done and using/re-using them for reference, performance, accountability, and evidence, and for economic and historical reasons.
Activity 6	*Process*	Using and manipulating records (and/or archives) within case management in business processes for reference, performance, accountability, evidence, and/or economic reasons.
Activity 7	*Distribute*	Distributing records for use within organizations.
Activity 8	*Structure*	Adding relevant structures to records (and/or archives) that help users in quickly finding and identifying them.
Activity 9	*Publish*	The external and/or internal publication of records (and/or archives), according to procedures and legal obligations.
Activity 10	*Analyse*	Analysing records (and/or archives) for knowledge gathering or management decisions based on defined or random queries or analysing tools using various (defined or random) algorithms
Activity 11	*Use/re-use*	Using and re-using records (and/or archives) for reference, performance, accountability, and evidence, and for economic and historical reasons.
Information Archiving		Archiving records (and/or archives) based on the four dimensions of information, the two archival principles, and the five requirements of information access.
Activity 12	*Contextualize*	Continuously adding new metadata to capture changes in situational and environmental contexts.
Activity 13	*Appraise*	Defining the relevance of records (and/or archives).
Activity 14	*Select*	Selecting records (to retain or to destroy).
Activity 15	*Retain*	Retaining records until the end of their retention period or indefinitely.
Activity 16	*Dispose*	Destroying records that have lost their relevance at the end of their retention period.
Activity 17	*Preserve*	Using preservation tools and techniques to retain records (and/or archives) indefinitely (or for a very long time).
Activity 18	*Secure*	Using information security measures and technologies to secure records (and/or archives).
Information Auditing		Auditing records (and/or archives) according to the four dimensions of information, the two archival principles, and the five requirements of information access.
Activity 19	*Audit*	Audit records (and/or archives) according to arranged requirements.

Table 1. Primary processes of the information value chain and their activities

information value chain, three models were crucial: [1] the model of the knowledge value chain of Holsapple and Singh (2001), [2] the recordkeeping model of Peter Horsman (1999, 2001), and [3] the InterPARES Chain of Preservation (Jansen 2015).

The value chain allows EIM to:
1. provide proper control of the performance of business processes;
2. provide trusted information;
3. assist in the realisation of the governance and compliance efforts of organizations;
4. provide legal readiness;
5. provide in the protection of sensitive records; and
6. assist in the construction of trusted archives.

The information value chain can be used by EIM to identify possible risks for the organization and to take proper actions if breaches of laws and regulations take place (Bearman, 2006; Van de Pas and Van Bussel, 2015ab). Tables 1 and 2 give an overview of the information value chain.

Information Leadership	Establishing management conditions, ethics, and circumstances that enable and facilitate EIM.
Information Coordination	Managing dependencies to ensure that EIM processes and resources are used adequately at appropriate times.
Information Control	Ensuring that information professionals and resources are available in sufficient quantity and quality, of course subject to security requirements.
Information Measurement	Assessing values of resources, information professionals, and their deployment.
Information Maintenance	Ensuring that the original condition of assets or resources within the information infrastructure are conserved as nearly, and as long, as possible, are compensated for normal wear and tear, and are renewed when necessary.

Table 2. Secondary processes of the information value chain

4.5. The behavioural component of the 'Archive-as-Is': Organizational Behaviour (E)

From a psychoanalytical point of view, Ihanus (2007) recognizes three phases of archival registrations: archivalization, archivization, and archiving. *Archivalization* has been defined by Eric Ketelaar (2000a, p. 329; 2001, p. 132-133)) as 'the conscious or unconscious choice (determined by social and cultural factors) to consider something worth archiving'. Ketelaar refers to the social psychologist Geert Hofstede (1997, p. 5), who defined 'culture' as 'the software of the mind', the 'collective programming of the mind which distinguishes the members of one group or category of people from another'. Humans do have, according to Hofstede, the ability to deviate from this programming, but it is clear that it affects the way employees are acting and thinking in business processes. This mental programming affects the way people intuitively consider something 'worth keeping' – or not. After archivalization, a more conscious choice is made about *archivization* (in the

Derridean sense (see Part I of this article)), about externalizing archivalization's choice in inscribing a trace in an external location. The last, conscious phase is *Archiving*, capturing and filing a record into the (organizational) archive. Between these three phases are psychological filters, and interplays between unconsciousness and consciousness. The first two phases of registrations determine whether (and how) actions are externalized and inscribed in archives. They determine the way people behave. They define behaviour that influences the way people construct, process, and use archives and the way archivists acquire, contextualize, and appraise archives and records. Ketelaar assumes that people working within the same organization will use and create records in different ways (Ketelaar, 2000a, p. 328).[1]

Different organizations are implementing the information value chain differently. Professional standards lead to different ways of creating and using records and archives. For understanding records and archives, employees and archivists of organizations are to be known in their social, religious, cultural, political, and economic contexts (Ketelaar, 2000a, 2001). These contexts define the 'software of the mind', and the effects of human behaviour that are its consequences. The 'software of the mind' impresses the fact that archives are not neutral, not complete, and a result of human behaviour within organizations. That behaviour reflects morals, preconceptions, and the limitations of the social and cultural environment of employees, and offers only a distorted view of reality. Or, maybe better, they allow for the construction of realities, excluding, other realities as a result of archivalization and, later, appraisal and selection (Ihanus, 2007).

The information value chain is embedded and largely configured by this behavioural component of the theoretical framework. Behaviour can have detrimental effects on organizational and personal archives. Managing records and constructing archives is strongly dependent on the working of organizational systems of controls, the methods and instruments used to strengthen such controls, and the behaviour of employees when confronted with these systems, methods, and instruments. When entering an organization, an individual employee brings personal characteristics, a personal social, ideological, ethical, religious, and cultural background, and experiences from other organizations. Employees have their expectations, goals, and ambitions. Those can change when they are interconnecting with other employees when working and collaborating. This affects the organization itself, and the organizational morals and ethics agreed upon may change those of the individual employee, or the other way around. It may explain why some people choose to leave an organization and others elect to stay (Griffin and Moorhead, 2014: 4-5). Hofstede (1997) found that specific attitudes and behaviours of employees differed significantly because of the values and beliefs that characterized their environment. The ways employees are handling information, the choices they are making, and the way they are behaving when confronted with systems of (information) control are heaviliy affected by these values and beliefs.

Study of behaviour and culture has never been part of archival science. The first to connect behaviour and culture explicitly with records and archives management are Gillian Oliver and Fiorella Foscarini (2013). They use the viewpoint of information culture to 'tackle the people problem'. Based on an inadequate introduction of information culture, they try to use the Information Culture Framework[2] for

analysing and assessing recordkeeping behaviour and practices. Although it is a very courageous and interesting exploration, they, in my opinion, do not really succeed in the endeavor to 'tackle the people problem'. It is not really a practical guide and only offers superficial ideas for assessment techniques and training that cannot be used to develop behavioural change programs. More problematic is that their work is extensively based on work of archival scientists and cultural theorists, which probably accounts for irrelevant chapters on records continuum, information continuum, and record keeping informatics. But their work neglects very relevant work done on organizational behaviour and culture within organization studies, such as Weick (1979), Shein (1992), Kotter and Heskett (1992), Simon (1997), O'Donovan (2006), Robbins and Langton (2007), and many more.

The effects of behaviour in organizations on information and information management are already known for a very long time. Campbell (1958), Wilensky (1967), Downs (1967), Janis (1972), Kaufman (1973), Athanassiades (1973), O'Reilly (1978), and others, have provided considerable evidence of organizational dysfunctions attributed to failures in the information value chain. The hypothesis of Benjamin Singer (1980) was that organizations suffer from psychotic and pathological behaviours, just like people do, but are rarely diagnosed with it or treated as such. According to Singer (1980, p. 48), dysfunctional organizational behaviours often take the form of 'crazy systems' that generate 'confusion, error, and ambiguity' and even 'inscrutability and unaccountability, involving harm to the victim and often to the system itself, [breeding] a new kind of organizational trap' called Kafka circuits. These involve 'blind alleys, crazy situations', and processes that 'end where they began'. More recently, Ronald Rice and Stephen Cooper (2010) confirmed that information is often blocked or distorted in organizational communications. They state convincingly that organizations allow employees to (consciously or unconsciously) misuse, distort, or suppress information and records (Rice and Cooper, 2010, chapters 7 and 8). Zmud (1990) argued that the use of ICTs make organizational functions vulnerable to strategic information behaviours such as distortion of records. It is quite clear that employee behaviour can have detrimental effects of the way records are created, processed, managed, and communicated (Singer, 1980; Clegg et al, 2016).

Especially in bureaucratic organizations, information access might be (or will be) influenced by the intentional or unintentional choices employees make when handling records and when deciding which information to keep (or not). These

---

[1] Although the concept of archivalization is mentioned many times in archival literature, there is almost no research done on the concept since its introduction almost seventeen years ago. The concept is completely misrepresented in literature and is identified as (a step in) the appraisal of records and archives. But it is a psychological phenomenon that influences human behaviour. As such, it defines appraisal and selection, but it cannot be considered part of them. For an interesting study in which the concept is applied on archival institutions and social communities and in which some of its psychological nature is expressed: Mark A. Matienzo, *'Canonization, Archivalization, and the 'Archival Imaginary''*, Paper presented at Archive Fervour/ Archive Further: Literature, Archives, and Literary Archives, Aberystwyth, Wales, July 9-11, 2008. Online source. Archived at: http://hdl.handle.net/10150/216929 (retrieved on December 22, 2016).

[2] The name of their framework is not unique. The name has been used for completely different Information Culture Frameworks by M.N. Khan and F.T. Azmi (2005). 'Reinventing business organisations: the information culture framework'. *Singapore Management Review*, Vol, 27, No. 2, pp. 37-62, and Y. Zheng (2005). 'Information culture and development: Chinese experience of e-health', *Thirty-Eighth Annual Hawaii International Conference on System Sciences*, (Hicss 38). 2005. Big Island, Hawaii, Los Alamitos, California IEEE Computer Society, pp. 153a, 1-11.

choices affect logistics, access, quality, and context of records. Employee choices are influenced by many variables and reasons, among which 'power', resistance to overbearing control systems, and their specific individual background are extremely important ones.

The legal frameworks that are usually created to curb organizational misbehaviour, and the internal compliance processes that are implemented in many organizations are specifically designed to identify and punish those individuals and organizations that are implicated in misbehaviour. These measures attempt to dissuade organizations and their employees from engaging in misbehaviour by threatening to hold them to account for their actions and decisions. The prevalent views of the organization as a 'machine', characterized by stringent rules and procedures, standardization, centralization, task specialization, and ignoring relational (and social) dynamics, is (in scientific literature) increasingly being replaced by an awareness of the way relational dynamics within duties, responsibilities, and accountability requirements are developing (Painter-Morland, 2007ab). As Melvin Dubnick and H. George Frederickson (2011, p. 7-12) explain, accountability relationships are mostly in evidence *after an event* ('post factum'). These relationships include 'post factum' attempts to handle responsibility for human or organizational errors based on 'pre-factum' (*before an event*) expectations and assumptions on organizational behaviour, configured in ICTs. Such attempts are largely based on records about these facts. ICTs are defined and configured 'pre-factum' and reflect expectations and assumptions of behaviour, but they do not reflect the activities and behaviour 'per factum', *during an event* (Heidelberg, 2015). Neglected also are how these activities relate to decisions within real actions and transactions and how to be accountable for those 'per factum' activities (Heidelberg, 2015). Organizations try to eliminate this stage by designing ICTs to avoid social relations, to avoid political discussion and debate, and to avoid infringement on rules. But these rules are only strengthening the bureaucratic system itself. That system hides 'spaces of contestation', spaces that should be filled with possibilities for political discussions, negotiations, and debates, where decisions are reached and where accountability should be prominent. Heidelbergs 'per factum' theory stresses the situated, relational dynamics *during* current policies, decisions, actions, and transactions (Heidelberg, 2015, p. 10, 18). But because 'per factum' is neglected (or, maybe, denied), these activities are not or only marginally captured in records and archival systems, therefore prohibiting records to document the most important spaces of decision making within organizations.

The activity theory can be used to explain the conflicts that exist within such a mechanistic view on information processing (Kaptelinin and Nardi, 2012). Bonnie Nardi (1996, p. 5) argued that mediation is a core concept of activity theory: human experience is shaped by the tools and sign systems in use. Nardi (1996, p. 7-13) emphasizes the importance of motive and consciousness, which are human characteristics that differentiate between people and things. People are not 'nodes' or 'agents' in a system. They are actors using systems as a tool to realize objectives. People and machines process information different. They cannot be modelled in the same way. When that happens, deviant human behaviour will be a result. This explains why configuring systems 'pre factum' to avoid 'per factum' is not going to

work. It is not the way humans process information. EIM will need to address human behaviour in a way that allows employees to use ICTs as a tool that allows for relational dynamics in the 'per factum' stage.

The accountability metaphors of the Agora and the Bazaar, proposed by Ciarán O'Kelly and Melvin Dubnick (2015) to characterize neglected 'spaces of contestation', stress the importance of relational dynamics in the 'per factum' stage. An Agora is a social environment in which purposes, reasons, and norms are developed. It is 'a fluid, contingent and localised accountability space, founded on an unending cascade of social situations and relationships' in and between organizations (O'Kelley and Dubnick, 2015, p. 9). This space is linked to collaboration between participants in that space, based on norms that focus on the fairness of aims and procedures. Organizational procedures, managerial power structures, and organizational purposes are 'informed' about the standpoints and decisions emerging as results from these collaborative relationships. These results are developed within a context where people combine moral sentiments with ethical requirements and constraints. The defining reasons for action are generated 'per factum'. Within the Agora, the metaphor of the Bazaar describes exchange in mutual pursuit of *each other's interests*. The focus of the Bazaar is on the negotiations that generate results, and the exchanges needed for those negotiations to be successful. These exchanges assist people in developing standpoints and decisions, trying to find a mutual interest and willing to trade favours, information, or esteem to achieve their purposes (O'Kelly and Dubnick, 2015, p. 9-16). It is what is happening in daily organizational practice. Exchanges within the Bazaar may not be open to description, formal scrutiny, codified rules, or bureaucratic control. Such spaces may involve psychotic and pathological behaviours (Singer, 1980).

It is spaces like the Agora and the Bazaar where the effects of organizational behaviour are extremely dominant. Much of the exchanges in the Agora and the Bazaar are not captured in records until after the event ('post factum'), according to 'pre-factum' procedures, and, thus, distorting social and situational reality. It will be EIM's challenge to organize the information value chain in such a way that employees use flexible ICTs within spaces of contestation for that will not only allow accountability to be based more on relational dynamics but it will also guide human behaviour to focus on organizational objectives.

5. Concluding remarks and future research

I started this research with the objective of finding a viable theoretical foundation for EIM. This foundation should allow EIM to capture unstructured information objects into its management procedures, tools, and ICTs to end the existing 'information chaos' and to improve the organizational ability to reach business objectives and to define business strategies. The concepts of records and archives are crucial for those endeavors. In the first part of this article, I showed that only within archival science theoretical frameworks have been developed using those concepts. Those theories offer valuable contributions and insights for EIM, but do not have a focus on reaching organizational objectives or defining business strategies. Their focus is on cultural (or historical) value and evidential value, not on the

organizations that are generating records and constructing archives. In the second part of this article, I defined the theoretical framework of the 'Archive-as-Is' that emphasizes the organizational value of the archive and the organizational challenges that EIM has to solve.

The theoretical framework of the 'Archive-as-Is' is primarily an *organizational* theory on records and archives. The focus of the framework is on the organizations (and/or persons) that construct archives and create, process, manage, and preserve records in their business processes and activities. The framework is based on the philosophy of pragmatism. As is common with each pragmatic theory, there is an unmistakable relationship with organizational practice. This relationship expresses itself in the framework's components that are all directly influencing organizational policies, business processes, actions, and transactions. They have to be recognized by EIM to improve the organizational processing of records and archives, to fight 'information chaos', and to guide organizational behaviour.

In the archival spectrum, the framework finds its place between the *context oriented theory* of the Records Continuum and the *records oriented theory* of Digital Diplomatics. Both theories have influenced the framework. It may be called an *organization oriented archival theory*. That is an orientation that is just as indispensable in a digital world as the context and object orientations are. It has been 'forgotten' in the frenzy of exciting research following the 'archival turn'. The framework is a declarative model for understanding the archive 'as is', how it has been designed, constructed, processed, manipulated, and managed, and how it has 'grown' to be the archive that the organization or the person that generated it, wants it to be. The three defining components of the theoretical framework can be used by EIM as an analytical tool to ascertain if all conditions for managing records and archives are met. The fourth component, the information value chain, offers a model for EIM to define and implement primary and secondary processes (and related activities) to realize the dimensions of information, the archival principles, and the requirements of information access. Organizational behaviour, the fifth component, stresses the necessity for EIM (and for archivists) to contextualize organizational practice, to allow for flexible ICTs to offer employees the possibility to use spaces of contestation 'per factum', and to be prepared for distortion of archives.

Archives shape and control the way history is read. They do. But archives are, from the moment of their construction, distortions of reality, leading to biased images of the past. Contextualizing will be crucial to 'correct' that distortion as much as is possible although the simplified metadata that capture context will also be distorting reality. In the end, the archive is as it is, a construct configured, managed, and preserved according to organizational (or personal) demands and desires, with gaps as a result of appraisal and selection, and, as a consequence, presenting a social reality that is only mirroring a very simplified and distorted view of the contexts in wich the records and the archive were generated.

Further research is an absolute necessity. It is necessary to see if the theory can be used as an analytical tool for EIM. The relationships between the components of the framework need to be studied more in depth. The relationship between EIM, the

theoretical framework, and the realization of organizational objectives needs more research. Research is necessary to see if Muller, Feith, and Fruin's statement about 'organically grown archives' is correct within digital environments. The effects of organizational behaviour on records and archives in daily organizational practice are neglected in archival research projects at the moment, although they are crucial to explain why the archive is as it is. I think the biggest challenge for EIM is to find ways to guide organizational behaviour in constructing and contextualizing archives. More research is needed in organizational behaviour and human-computer interaction within spaces of contestation that extremely influence accountability and archiving. Activity theory may be a very useful theory for research in that regard.

Literature

Acland, G. (1991). Archivist-Keeper, Undertaker, or Auditor. *Archives and Manuscripts, 19*(1), 13-14.

Arnold, J., with J. Silvester, F. Patterson, I. Robertson, C. Cooper, & B. Burnes (2005). *Work Psychology: Understanding Human Behaviour in the Workplace (*4th ed.). Harlow: Pearson Education Ltd.

Athanassiades, J.C. (1973). The distortion of upward communication in hierarchical organizations. *The Academy of Management Journal, 16*(2), 207-226.

Baker, M. (2013). *Findability is a content problem, not a search problem.* (May 28). Retrieved from http://everypageispageone.com/2013/05/28/findability-is-a-content-problem-not-a-search-problem/. Archived at http://web.archive.org/web/20160405191132/ http://everypageispageone.com/2013/05/28/findability-is-a-content-problem-not-a-search-problem/.

Barwise, J., & J. Perry (1983). *Situations and Attitudes*, Cambridge, MA: MIT Press.

Bearman, D. (1993a). Record Keeping Systems. *Archivaria, 36*, 16-36.

Bearman, D. (2006). Moments of risk. Identifying threats to electronic records. *Archivaria, 62*, 15-46.

Berlin, L.M., R. Jeffries, V.L. O'Day, A. Paepcke, & C. Wharton (1993). Where did you put it? Issues in the design and use of a group memory. In B. Arnold, G. Van der Veer, T. White (eds.), *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems* (pp. 23-30). New York: ACM.

Boudrez, F., H. Dekeyser, & J. Dumortier (2005). *Digital Archiving. The new challenge.* Mont Saint Guibert: IRIS.

Bruschi, C. (2009). *The Wandering Heretics of Languedoc*. Cambridge (UK): Cambridge University Press.

Buckland, M.K. (1991). *Information and information systems*. Westport: Greenwood Publishing Group.

Buneman, P., S. Khanna, & W.C. Tan (2001). Why and where. A characterization of data provenance. In J. Van den Bussche, V. Vianu (eds.), *Database Theory. ICDT-2001. Proceedings of the 8th International Conference on Database Theory, London, January 4-6* (pp. 316-330). Berlin / Heidelberg: Springer.

Burnett, G., P.T. Jaeger, & K.M. Thompson (2008). Normative behavior and information: the social aspects of information access. *Library & Information Science Research, 30*(1), 56-66.

Campbell, D.T. (1958). Systematic error on the part of human links in communication systems. *Information and Control, I*, 334-369.

Clegg, S., M. Pina e Cunha, I. Munro, A. Rego, & M. Oom de Sousa (2016). Kafkaesque power and bureaucracy. *Journal of Political Power, 9*(2), 157-181.

Cook, T. (1997). What is past is prologue. A history of archival ideas since 1898, and the future paradigm shift. *Archivaria, 43*, 17-63.

Cook, T. (2007). Electronic Records, Paper Minds: the revolution in information management and archives in the post-custodial and post-modernist era. *Archives & Social Studies: A Journal of Interdisciplinary Research, 1,* 399-443. (reprint from 1994).

Cox, R.J., & H.W. Samuels (1988). The archivist's first responsibility. A research agenda to improve the identification and retention of records of enduring value. *The American Archivist, 51*(Winter/Spring), 28-42.

Cui, Y., & J. Widom (2003). Lineage tracing for general data warehouse transformations. *The VLDB Journal - The International Journal on Very Large Data Bases, 12*(1), 41-58.

Cunningham, A. (2015). Postcustodialism. In L. Duranti, P.C. Franks (eds.), *Encyclopedia of Archival Science* (pp. 274-278). London: Rowman and Littlefield.

Dervin, B. (1997). Given a context by any other name. Methodological tools for taming the unruly beast. In P. Vakkari, R. Savolainen, B. Dervin (eds.), *Information seeking in context* (pp. 13-38). London: Taylor Graham.

Dervin, B. (2003). From, the minds eye of the user. The Sense-Making Qualitative-Quantitative methodology. In B. Dervin and L. Foreman-Wernet (eds.), *Sense-Making Methodology Reader*. New York: Hampton Press. (First published in 1992.)

Devlin, K.J. (1991). Oracles in situation semantics. In J. Barwise, J.M. Gawron, G. Plotkin, and S. Tutiya (eds.), *Situation Theory and its Applications, Vol. 2* (Chapter 3, pp. 41-49). Menlo Park: CLSI.

Devlin, K. (1994). Situation Theory and Social Structure. In M. Masuch and L. Polos (eds.), *Knowledge Representation and Reasoning under Uncertainty* (pp. 197-237). Berlin: Springer-Verlag.

Devlin, K., & D. Rosenberg (2008). Information in the study of human interaction. In J. Van Benthem, P. Adriaans, D.M. Gabbay, P. Thagard, and J. Woods (eds.), *Handbook of the Philosophy of Information* (pp. 685-710). Amsterdam: Elsevier.

Dewey, J., & J. Tufts (1908). *Ethics*. New York: Henry Holt and Co.

Dollar, C. (1992). *Archival theory and information tehnologies. The impact of information technologies on archival principles and methods*. Macerata: University of Macerata.

Downs, A. (1967). *Inside Bureaucracy.* Boston: Little-Brown.

Dubnick, M.J., & H.G. Frederickson (2011). *Public accountability. Performance measurement, the extended state, and the search for trust.* Washington, DC: Kettering Foundation & National Academy of Public Administration.

Duranti, L. (1997a). The preservation of the integrity of electronic records. In *Proceedings of the DLM-Forum on electronic records. Brussels, 18-20 December 1996* (pp 60-65). Luxemburg: Office for Official Publications of the European Communities.

Duranti, L. (1997b). The Archival Bond. *Archives & Museum Informatics, 11*(3-4), 213-218.

Duranti, L. (2007). Archives as a Place. *Archives & Social Studies: A Journal of Interdisciplinary Research, 1*(0), 445-466. (Reprint of 1996).

El Kharbili, M., S. Stein, I. Markovic, & E. Pulvermüller (2008). Towards a framework for semantic business process compliance management. In S. Sadiq, M. Indulska, and M. zur Muehlen (eds.), *Proceedings of the Workshop on Governance, Risk and Compliance for Information Systems (GRCIS 2008), CEUR, Workshop Proceedings, Vol. 339, Montpellier* (pp. 1-15). Retrieved from http://ceur-ws.org/Vol-339/

Frege, G. (1980). *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number* (translated by J.L. Austin, 2nd ed.). Chicago: Northwestern University Press.

Gilliland, A.J., McKemmish, S., & Lau, A.J. (eds.) (2016). *Research in the Archival Multiverse*. Clayton, VIC: Monash University Publishing. http://dx.doi.org/10.26530/OAPEN_628143

Göpferich, S. (2006). Comprehensibility assessment using the Karlsruhe Comprehensibility Concept. *The Journal of Specialised Translation*, 6(11), 31-53.

Greetham, D. (1999). Who's in, who's out. The cultural politics of archival exclusion. *Studies in the Literary Imagination, 32*(1), 1-28.

Griffin, R.W., & G. Moorhead (2014). *Organizational behavior. Managing people and organizations* (11th ed.). Mason, OH: South-Western Cengage Learning.

Groth, P.T. (2007). *The origin of data. Enabling the determination of provenance in multi-institutional scientific systems through the documentation of processes*. Doctoral thesis at the University of Southampton. Retrieved from https://eprints.soton.ac.uk/264649/

Hartshorne, C., & P. Weiss (1933). *The Collected Papers of Charles Sanders Peirce*, IV. Boston: Harvard University Press.

Heidelberg, R.L. (2015). Political accountability and spaces of contestation. *Administration & Society*, April 14, 1-24.

Hill, H. (2013). Disability and accessibility in the library and information science literature: A content analysis. *Library & Information Science Research, 35*(2), 137-142.

Hockx-Yu, H. (2006). Digital Preservation in the Context of Institutional Repositories. *Program: Electronic Library & Information Systems, 40*(3), 232-243.

Hofstede, G. (1997). *Cultures and Organizations: Software of the Mind.* New York: McGraw-Hill. (Most recent (third) edition: 2010).

Holsapple, C.W., & M. Singh (2001). The knowledge chain model: activities for competitiveness. *Expert Systems with Applications, 20*(1), 77-98.

Horsman, P.J. (1994). Taming the elephant. An orthodox approach to the principle of provenance. In K. Abukhanfusa, J. Sydbeck (eds.), *The principle of provenance. First Stockholm Conference on archival theory and the principal of provenance, 2-3 september 1993, Stockholm* (pp. 51-63).

Horsman, P.J. (1999). Archiefsystemen en kwaliteit. In P.J. Horsman, F.C.J. Ketelaar, T.H.P.M. Thomassen (eds.), *Naar een nieuw paradigma in de archivistiek.* Den Haag: SAP.

Horsman, P.J. (2001). *Electronic Recordkeeping. The Recordkeeping System as framework for the management of electronic records.* Amsterdam; Archiefschool.

Horsman, P.J., F.C.J. Ketelaar, & T.H.P.M. Thomassen (1998). *Tekst en context van de Handleiding voor het ordenen en beschrijven van archieven van 1898.* Hilversum: Verloren.

Hurley, C. (1995). Ambient functions: abandoned children to zoos. *Archivaria, 40*(Fall), 21-39.

Hurley, C. (2005). Parallel provenance [Series of parts]: Part 1: What, if anything, is archival description? *Archives and Manuscripts, 33*(1), 110-145.

Ihanus, J. (2007). The archive and psychoanalysis: Memories and histories toward futures. *International Forum of Psychoanalysis. 16*(2), 119-131.

Israel, D., & J. Perry (1991). Information and architecture. In J. Barwise, J.M. Gawron, G. Plotkin, and S. Tutiya (eds.), *Situation theory and its applications Vol. 2, CSLI Lecture Notes 26* (pp. 147-160). Stanford, CA: Center for the Study of Language and Information (CSLI).

James, W. (1907). *Pragmatism: A New Name for Some Old Ways of Thinking*. Cambridge, MA: The Riverside Press.

James, W. (1909). *The meaning of truth*. Cambridge, MA: The Riverside Press.

Janis, I. (1972). *Victims of Groupthink: A psychological study of foreign-policy decisions and fiascoes*. Boston: Hughton-Mifflin.

Jansen, A. (2015). Chain of Preservation. In L. Duranti, and P.C. Franks (eds.), *Encyclopedia of Archival Science* (pp. 133-136). London: Rowman and Littlefield.

Jenkinson, H. (2003). Modern Archives: Some Reflections on T. R. Schellenberg: Modern Archives. Principles and Techniques. In R.H. Ellis, P. Walne (eds.), *Selected writings of Sir Hilary Jenkinson* (pp. 339-342). Chicago: SAA.

Jones, W. (2011). No knowledge but through information. In D.J. Pauleen, G.E. Gorman (eds.), *Personal knowledge management: Individual, organizational and social perspectives* (pp. 143-166). Farnham: Gower Publishing Ltd.

Kaptelinin, V., & B. Nardi (2012). *Activity theory in HCI. Fundamentals and reflections*. Williston, VT: Morgan & Claypool.

Kaplan, E. (2000). We are what we collect, we collect what we are. Archives and the construction of identity. *The American Archivist, 63*, 126-151.

Kato, T., & M. Hori (2006). Beyond Perceivability. Critical requirements for universal design of information. In *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and accessibility* (pp. 287-288). Portland, OR: ACM.

Kaufman, H. (1973). *Administrative Feedback. Monitoring subordinates' behavior.* Washington DC: Brookings Institute.

Ketelaar, E. (2000a). Archivistics research saving the profession. *The American Archivist, 63*, 322-340.

Ketelaar, E. (2000b). De culturele context van archieven. In P.J. Horsman, F.C.J. Ketelaar, and T.H.P.M. Thomassen, *Context. Interpretatiekaders in de archivistiek* (pp. 83-91). Den Haag: SAP.

Ketelaar, E. (2001). Tacit narratives. The meaning of archives. *Archival Science, I,* 131-141.

Ketelaar, E. (2007). Archives in the Digital Age. New uses for an old science. *Archives & Social Studies, 1*(0), 167-191.

Kotter, J., & J.L. Heskett (1992). *Corporate Culture and Performance*. New York: Free Press.

Kozak, G. (2015). Access/Accessibility. In, L. Duranti, P.C. Franks (eds.), *Encyclopedia of Archival Science* (pp. 1-3). London: Rowman and Littlefield.

Kulthau, C.C. (2006). Kuhlthau's Information Search Process. In K.E. Fisher, S. Erdelez, L. McKechnie (eds.), *Theories of Information Behavior* (pp. 230-234). New Jersey: Information Today.

Lane, V., & J. Hill (2010). Where do we come from? What are we? Where are we going? Situating the archive and archivists. In J. Hill (ed.), *The Future of Archives and Recordkeeping. A reader* (pp. 3-22). London: Facet publishing.

Latour, B. (1990). Postmodern? No, simply amodern! Steps towards an anthropology of science. *Studies In History and Philosophy of Science, 21*(1), 145-171.

Le Roy Ladurie, E. (1975). *Montaillou, village occitan de 1294 à 1324*. Paris: Gallimard.

Lemieux, V.L., & the ImProvenance Group (2016). Provenance: Past, Present and Future in Interdisciplinary and Multidisciplinary Perspective. In V.L. Lemieux (ed.), *Building trust in Information. Perspectives on the Frontiers of Provenance* (pp. 3-45). Cham (ZG): Springer International Publishing AG.

Levy, D.M. (2001). *Scrolling forward. Making sense of documents in the digital age.* New York: Arcade Publishing.

Mason, R.O. (1986). Four ethical issues of the information age. *MIS Quarterly, 10*(March), 5-12.

Mathiesen, K. (2014). Facets of access: A conceptual and standard threats analysis. In *iConference 2014 Proceedings* (pp. 605-611). Berlin: iSchools.

Mayer-Schönberger, V. (2009). *Delete: The Virtue of Forgetting in the Digital Age*. Princeton / Oxford: Princeton University Press.

McCreadie, M., & R.E. Rice (1999). Trends in analyzing access to information. Part I. Cross-disciplinary conceptualizations of access. Part II. Unique and integrating conceptualizations. *Information Processing & Management, 35*(1), 45-76 and 77-99.

Michetti, G. (2016). Provenance: An archival perspective. In V.L. Lemieux (ed.), *Building trust in Information. Perspectives on the Frontiers of Provenance* (pp. 59-68). Cham (ZG): Springer International Publishing AG.

Muller, S., J.A. Feith, & R. Fruin (2003). *Manual for the arrangement and description of archives* (translated by A.H. Leavitt, and with new introductions by P. Horsman, E. Ketelaar, T. Thomassen and M.R. Barit). Chicago: SAA.

Narayan, B., & M. Olsson (2013). Sense making across space and time. Implications for the organization and findability of information. In F. Bouthillier, B. Yu, A. Grove (eds.), *Proceedings of the 76th ASIS&T Annual Meeting: Beyond the Cloud: Rethinking Information Boundaries* (Article 72, pp. 1-9). Silver Springs, MD: American Society for Information Science.

Nardi, B.A. (1996). Activity theory and human computer interaction. In B.A. Nardi (ed.), *Context and Consciousness: Activity Theory and Human-Computer Interaction* (pp. 7-16). Cambridge MA: The MIT Press.

Nardi, B., & V.L. O'Day (1999). *Information Ecologies. Using technology with heart*. Cambridge / London: The MIT Press.

Nesmith, T. (1999). Still fuzzy, but more accurate. Some thoughts on the 'ghosts' of archival theory. *Archivaria, 47*, 136-150.

Nesmith, T. (2015). Principle of Provenance. In L. Duranti, P.C. Franks (eds.), *Encyclopedia of Archival Science* (pp. 284-288). London: Rowman and Littlefield.

OECD (2015). OECD Skills Outlook 2015. *Youth, skills and employability*. Paris: OECD.

O'Donovan, G. (2006). *The Corporate Culture Handbook. How to plan, implement and measure a successful culture change programme*, Dublin, The Liffey Press.

O'Kelley, C., & M. Dubnick (2015). *Accountability and its metaphors. From forum to agora and bazaar (Paper presented to the PSG VII track (Quality and Integrity of Governance) of the 2015 EGPA Annual Conference August 24-29, 2015, Toulouse, France, Toulouse, EGPA)*. Retrieved from http://pure.qub.ac.uk/portal/files/13032528/COK_MJD_EGPA_Paper.pdf.

O'Reilly, C.A. (1978). The intentional distortion of information in organizational communication: A laboratory and field investigation. *Human Relations, 31*(2), 173-193.

Oliver, G., & F. Foscarini (2013). *Records Management and Information Culture. Tackling the people problem*. London: Facet Publishing.

Painter-Morland, M. (2007a). Defining Accountability in a network society. *Business Ethics Quarterly, 17*, 515-534.

Painter-Morland, M. (2007b). Redefining accountability as relational responsiveness. *Journal of Business Ethics, 16*, 89-98.

Penco, C. (1999). Objective and cognitive context. In P. Bouquet, L. Serafini, P. Brezillon, M. Benerecetti, and F. Castellani, (eds.), *Modeling and using contexts. Proceedings of the second international and interdisciplinary conference Context '99* (pp. 270-283). Berlin / Heidelberg: Springer.

Pettigrew, A.M. (1979). On studying organizational cultures. *Administrative Science Quarterly, 24*(4), 570-581.

Pettigrew, A.M. (1990). Longitudinal Field Research on Change: Theory and Practice. *Organization Science, 1*(3), 267-292.

Pfeffer, J., & G.R. Salancik (1978). *The external control of organizations. A resource dependence perspective*. New York: Harper & Row.

PIVOT (1994). *Handelend optreden. Overheidshandelen: modellen, onderzoeksmethoden en toepassingen*. Den Haag: PIVOT.

Porter, M., & V.E. Miller (1985). How information gives you competitive advantage. *Harvard Business Review, 63*(4), 149-160.

Posner, E. (1967). Max Lehmann and the genesis of the principle of provenance. In E. Posner, *Archives and the public interest. Selected essays* (pp. 135-144). Washington: Society of American Archivists.

Puri, C., D.S. Kim, P.Z. Yeh, & K. Verma (2012). Implementing a data lineage tracker. In A. Cuzzocrea and U. Dayal (eds.), *Data Warehousing and Knowledge Discovery. Proceedings of the 14th International Conference, DaWaK 2012, september 3-6 2012, Vienna* (pp. 390-403). Springer.

Reilly, T. (2005). *From provenance to practice. Archival theory and 'return to community' (paper presented at the First Nations, first thoughts Conference, University of Edinburgh.* Retrieved from http://prism.ucalgary.ca/bitstream/1880/47398/1/Reilly_From_Provenance.pdf.

Resmini, A., & L. Rosati, From physical to digital environments (and back). Seven laws of findability. In *Translating Information Architecture: Proceedings of Europe's third Information Architecture summit (EuroIA)* (pp. 162-170). Barcelona: ASIS&T.

Rice, R.E., & S.D. Cooper (2010). *Organizations and Unusual Routines. A Systems Analysis of Dysfunctional Feedback Processes*. Cambridge (UK): Cambridge University Press.

Robbins, S.P., & N. Langton (2007). *Organizational Behaviour: Concepts, Controversies, Applications* (4th ed.). North York (Ontario): Pearson Education Canada.

Saracevic, T. (2007a). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: nature and manifestations of relevance. *Journal of the American Society for Information Science and Technology, 58*(3), 1915-1933.

Saracevic, T. (2007b). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effects of relevance. *Journal of the American Society for Information Science and Technology, 58*(13), 2126-2144.

Schellenberg, T. (2003). *Modern Archives. Principles and techniques*. Chicago: SAA.

Self, D., A. Armenakis, & M. Schraeder (2007). Organizational change content, process, and context. A simultaneous analysis of employee reactions. *Journal of Change Management, 7*(2), 211-229.

Shein, E. (1992). *Organizational Culture and Leadership: A Dynamic View*. San Francisco: Jossey-Bass.

Shields, P. (1998). Pragmatism as philosophy of science. A tool for public administration. In J.D. White (ed.), *Research in Public Administration Vol. 4* (pp. 195-226). Bingley: Emerals Group Publishing.

Simmhan, Y., B. Plale, & S. Gannon (2005). A survey of data provenance in e-science. *ACM SIGMOD, 34*(3), 31-36.

Simon, H.A. (1997). *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organizations* (4th ed.). New York: The Free Press.

Singer, B.D. (1980). Crazy Systems and Kafka Circuits. *Social Policy, 11*, 46-54.

Stupariu, I. (2015). Defining the right to be forgotten. *A comparative analysis between the EU and the US*. Budapest: Central European University. (LL.M. short thesis).

Sweeney, S. (2008). The ambiguous origins of the archival prinicple of provenance. *Libraries & the cultural record, 43*(2), 193-213.

Thomassen, T.H.P.M. (1999). Paradigmatische veranderingen in de archiefwetenschap. In P.J. Horsman, F.C.J. Ketelaar, T.H.P.M. Thomassen (eds.), *Naar een nieuw paradigma in de archivistiek* (pp. 69-79). Den Haag: Stichting Archiefpublicaties.

Thomassen, T., B. Looper, & J. Kloosterman (eds.) (2001). *Toegang. Ontwikkelingen in de ontsluiting van archieven*. Den Haag: Stichting Archiefpublicaties.

Upward, F. (2017). The archival multiverse and eddies in the spacetime continuum. In A.J. Gilliland, S. McKemmish, A.J. Lau (eds.), *Research in the Archival Multiverse* (pp. 198-227). Clayton, VIC: Monash University Publishing. http://dx.doi.org/10.26530/OAPEN_628143

Upward, F., & S. McKemmish (1994). Somewhere beyond custody. *Archives and Manuscripts, 22*(1), 136-149.

Van Bussel, G.J. (2012a). *Archiving should be just like an Apple, en acht andere (nuttige?) stellingen*. Amsterdam: Amsterdam University Press.

Van Bussel, G.J. (2012b). Reconstructing the Past for Organizational Accountability. *The Electronic Journal of Information Systems Evaluation, 15*(1), 127 -137.

Van Bussel, G.J. (2016). An Accountability Challenge. Capturing records and their context in Enterprise Information Systems. In P. Silva, A. Guerreiro and R. Quaresma (eds.), *Proceedings of the 10th European Conference of Information Systems Management. ECISM 2016, Evora, Portugal, 8-9 September 2016* (pp. 204-211). Reading: ACPI.

Van Bussel, G.J., & F.F.M. Ector (2009). *Op zoek naar de herinnering. Verantwoordingssystemen, content-intensieve organisaties en performance*. Helmond: Van Bussel Document Services.

Van Bussel, G.J., & H. Henseler (2013). Digital Archiving and eDiscovery. Delivering evidence in an age of overload. In B. John, M. Nkhoma and N. Leung (eds.), *Proceedings of the 4th International Conference on Information Systems Management and Evaluation. ICIME 2013, Ho Chi Min City, Vietnam, 13-14 May 2013* (pp. 281-288). Reading.

Van de Pas, J., & G.J. van Bussel (2015a). Privacy lost – and found? The information value chain as a model to meet citizens' concerns. *Electronic Journal of Information Systems Evaluation, 18*(2), 199-209.

Van de Pas, J., & G.J. van Bussel (2015b). Embedding Privacy in ICT Architectures. The citizen as public stakeholder in architecture development. In B. van der Sloot (ed.), *Proceedings of the Amsterdam Privacy Conference (21-26 October 2015)* (14 pages, incl. references (only available on USB)). Amsterdam: APPR. Retrieved from http://www.vbds.nl/wp-content/uploads/2015/10/Van-de-Pas_-Van-Bussel.pdf.

Weick, K. (1979). *The Social Psychology of Organizing*. New York: McGraw-Hill.

Weick, K. (1995). *Sensemaking in Organisations*. London: Sage.

Wilensky, H. (1967). *Organizational Intelligence. Knowledge and Policy in Government and Industry*. New York: Free Press.

Wittgenstein, L. (1961). *Tractatus Logico-Philosophicus* (translated by D.F. Pears and B.F. McGuinness, first published in 1922). London: Routledge and Kegan Paul.

Zalamea, F. (2003). Peirce's logic of continuity. Existential graphs and non-Cantorian Continuum. *The Review of Modern Logic, 9*(29), 115-162.

Zmud, R.W. (1990). Opportunities for Strategic Information Manipulation through new information technology. In J. Fulk, C.W. Steinfield (eds.), *Organizations and Communication Technology (Chapter 5, pp. 95-116)*. London / New Delhi: Sage Publications.

RIENK JONKER

A perfect match?
Connecting partners in the
labyrinth of information[*]

Introduction

In this essay I will give some insight into the results of my quest as an archivist[1] to understand the fundamentals of information. It is not only about trying to understand the complexity, but mostly to grasp or grock this phenomenon. With the verb grock one indicates that he or she is to understand completely and thoroughly an object, subject or issue (Heinlein, 1961). At one point I was surprised to notice that the common perception of complexity of information and information management overshadows the relative simplicity of the outcome of the grocking.

As an archivist in Leeuwarden I am on an almost regular basis confronted with born digital information. I have to answer questions related to (functional) requirements and tenders for digital systems, change management, information architecture, system management, process development and management, migrations, and conversions. The answers have to be simple, preferably fitting on one page (e.g. as a checklist), because those who are asking are not interested in long considerations, reports and notes. Often there is not much time and quick decisions have to be made. It makes my job interesting, because old and new theories clash sometimes.

A couple of years ago I became aware that I had been using an implicit sort of internal reference model for a long time. It was based on what I learned during my work as an archivist and my work in ICT. In particular, the fundamentals of arranging and describing archives, the knowledge about describing archives that I acquired through the years and the rise of digital systems for this purpose were invaluable to me to understand today's digital development. In fact, this model was at its core based on a kind of constant undercurrent. This awareness was prompted to me through remarks that my colleague Frans Smit made about the behaviour of people working in ICT, and their internal reference frameworks. As a result, I could explicate my implicit model and turn the concepts into checklists and forms that are now in use in my organisation. In addition, I understood that from certain

---

[1] When I use the word archivist, both the archivist and records manager are included. The archivist as keeper of archives in archival holdings and the records manager as an official found in the frontline where the records are created, received and managed until their disposition.

perspectives and abstraction levels, age-old concepts such as filing structures and document types can be found in modern digital systems. In that sense, to paraphrase the writer of Ecclesiastes "there is nothing new under the sun". Because of this journey I could develop a basic information model or framework that at its core only consists of three basic interconnected elements. Each element has its own characteristics. For me, an information model is a somewhat formal abstract description of objects, attributes, relationships, and rules in a particular domain. It was a feasible answer to questions about the possibility to come up with an information model, including a model for sharing and exchanging information – independent of social, organisational and technical changes – and at the same time a model compliant to the required information qualities and to the interests to preserve the information. This model I defiantly call the Leeuwarder Information Model (LIM) because Leeuwarden is the place where it came to existence. Although it is a kind of archival information model, I found it somewhat insolent to call it that. There are enough other archival information models around.

Despite the organisational or cultural background of actors and despite what technical instruments are used for processing information, there is always a consistent undercurrent. An undercurrent from the viewpoint of information as a constant factor with basic interwoven notions or concepts as context, documents or more neutral information objects, activities with communication or exchange and particularly notions about their mutual dependencies.

The sum of these three elements, context, information object and activity can be called meaningful information (MI). The three elements also determine if a set of meaningful information makes up an archival record. Information is considered a record when it can be used in a personal or organisational context as evidence of a transaction, for reasons of compliancy, conformance and governance. It is mostly about administrative, legal and fiscal values. As a rule, one can say a record is always meaningful information, but not all meaningful information is a record. Meaningful information is only a record because and whenever we want it to be a record. One could say a record is a construct that lies in the eyes of the beholder. But to avoid further distraction, this essay is not about the definition of records and recordness. That is a discussion that will have to take place elsewhere.

A real surprise was to observe that essential parts of my notion of meaningful information could be annex to the concept of semantic information (SI) of Floridi (Floridi, 2010). This was a discovery I found worthwhile to look further into.

I will begin with an abstract point by introducing the concept of "something" which is at the base of my information model (LIM). Then I will outline an example of daily life which portraits what happens when an activity is triggered by an event: an activity starts, and information is processed. With that example, I can explain my information model and the concept of meaningful information. Then I will shortly point out the elements and characteristics that are part of the connection between the meaningful information of the LIM and the concept of semantic information of Floridi. This is preceded with some reflections about what influenced my quest.

The journey and the undercurrent

The world of information has all the characteristics of a chaotic complex system with lots of turbulence. It is in no way comparable to the familiar, almost static, paper environment. Nothing remains and everything changes, which is something we should accept as normal. We as archivists are in this fluid digital world looking for fixation points; we want to be able to fixate moments. But to be able to fix, it must be clear why something should be fixed, what content is to be fixed and eventually how this fixation is to be carried out technically. Our field, the archival science, has a broad variety of fundamental pillars at its disposal. To begin with – of course – the Manual for the Arrangement and Description of Archives (Muller, Feith, Fruin, 1898/1920). This work contains components and concepts that are still very useful in these modern times. This work has been supplemented by several useful standards from ICA, ISO, and even the European Union. Not to forget the results of the international research of InterPares (Duranti & Preston, 2008). The intention of these works is to provide the professional archivist guidance through the maze or the labyrinth of information concepts and information systems. But is this assumption of guidance right or are we missing something? If there is already enough guidance why is it that there are still so many questions asked, why is there so much unclear? In these turbulent times of digital change, archivists still need a sustainable fundamental layer on which they can build for a long time.

Perhaps the current concepts and models with derived terms and definitions are still insufficient for the approach of the current changes. A lot of those approaches are often accrued from a very persistent "paper mindset", a phenomenon and heritage which permeates our culture. It is a legacy that could be described as the paper paradigm. It is a phenomenon that leads to the 'paperisation' of the perception of the digital environment. This confusion must be resolved. Otherwise we will continue to drive forward at full speed, only relying on the rearview mirror while the windshield is still blinded. Therefore, a fundamental reinvention and redefinition of our professional paradigms is a prerequisite. On the other hand, the archivist as a practitioner has to travel on the moving high-speed train. And at the same time, he must find his way into design departments, construction sites, maintenance sites and boardroom table. This is only possible when the archivist's message is unambiguous and clear. To be able to do that he needs 'simple' information models with instruments that meet the following requirements: it is generic, easy to handle and constant.

It is becoming clearer and clearer that in the world of the infosphere, an archivist has to look far beyond the bubble of his own discipline. The infosphere is the whole informational environment constituted by all informational entities, their properties, interactions, processes, and mutual relations (Floridi, 2013). On the one hand, archivists need it to find extra support and partners in the complex new world, and on the other hand to make use of the recognised knowledge of those partners. But also, archivists must bring those partners into contact with archival science. It is a science with methodologies that is in a way a secure scientific theory that cannot be proven wrong, because it has always been reinvented in totally different contexts, environments, places and times – only the re-inventors are not aware of this fact.

Starting point

The starting point – the basis for this model – is the concept of the "something". With this concept it is possible, at a very high level of abstraction, to reveal the basic elements of an event or activity with its additional determining characteristics.

Each event or activity has an element of time; there is always a starting point and an endpoint. Both the beginning and the end are fixed (as a rule). It is also possible that an event has started or ended unnoticed. In that case a begin or an end can be defined on the go. In addition, associated with it there are one or more actors in human or other form like computers or machines. The "something" delivers, whether or not there will be a result. Also within an event or activity information is received, processed and sent.
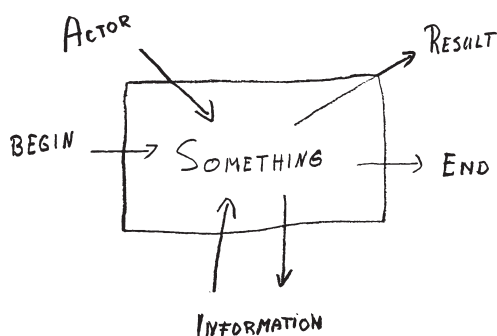
Figure 1. The "something"

Because of the starting and endpoint and because of the steps taken in between (including the expected results), one could say this "something" behaves like a process. On a certain level of abstraction, one could even say that every activity behaves like a process, whether it is a real process, a project or another kind of action and can be approached like an object or an entity.

In its essence the "something" is a representation of a constant undercurrent and can be described as an object. As undercurrent it is generic and, and it is also possible to show the minimal imperative components which make a "something" up, and in a generic way to describe its components like context, information objects and the activity. The process of adding metadata to those objects is implicitly executed by everyone in one way or another, but it is for sure recognisable for an archivist, because description is a very substantial part of his profession.

This "something" is for me a starting point of view to answer questions about information- and records management in my organisation. Because everything behaves like a process, for me also organisations, organisational entities, board governors, management boards even individual members of staff fit in this generic concept. Although sometimes there is a great difference in nature they are all temporarily because the start or begin at some moment and will also at some given time stop or expire.

An everyday example

The concept of the "something" can be outlined by using an example outside the comfort zone of an archivist. The example is about what I call the activity of refuelling a car. Imagine you are driving your car and on a given moment the

dashboard signals that the fuel tank is almost empty. It must be filled immediately, otherwise you will end up on the side of the road in the middle of nowhere. The signal is clear, and it cannot be ignored either.

Figure 2. Information object (gauge)

Figure 3. Activity - (re)fuelling a car

To stay out of trouble, the wisest thing to do is to go to the closest petrol station. If time is available you could also choose to go to your favourite station, or the cheapest station, it is your personal choice. Before you can start refuelling, some things must be settled first. For example, you must park the car at a pump that has the right kind of fuel. You also want to find the tank opening on the right side of the pump. In this way, juggling the hose is not necessary.

Figure 4. Information object (display)

Once you have parked, you must execute the start of a financial transaction. The supplier wants to get paid. This transaction begins with your debit card. You'll have to give the vendor permission (through the console) to debit the due amount from your account after fuelling. Via the same console, you can indicate what kind of fuel you need. The console tells you which pump you must use. After removing the debit card, you can refill the tank. The pump display informs you about the progress of the fuelling in terms of volume (litres) and the amount of money due at that moment. This comes in handy for motorists who only like to pay for ten Euros or ten litres at a time. With the information on the display you know when to stop refuelling. In most cases it is simple; people fill up the whole tank. Most people pay more attention to the hose and the tank opening than to the display.

```
Tango Leeuwarden
Tijnjedijk 76
8963 AD Leeuwarden

Voor vragen:
www.tango.nl/contact of
070-3152500


Datum   24-07-2017 22:52

*
Pomp 4              EURO 95
Volume              26.05 ℓ
Prijs           € 1.469/ℓ
TOTAAL            € 38.27
*
BTW    21.00 %     € 6.64
Netto             € 31.63
MAESTRO               OD
    ***************6007
STAN              879342
Handelaarsid      326487
Host ref.nr 560000019341
Autor. code       6WHXBO
Goedgekeurd
AID      A0000000043060
AC       A8098888DDFBDFD4

Stationsnummer      6378
Ticketnummer  4535540329
51-146516-90876991

Kuwait Petroleum Ned. BV
BTW NL 001250590 B01
Tango. De slimste keuze
als je moet tanken.
```

Figure 5. Information object (paper)

As soon as the tank is full, refuelling stops. At that point, in my view, the display gives no information about the fuelling process, but only delivers information about the final condition. You can see the total price you must pay and the amount of fuel that is delivered. Once you put the hose back in its place, the payment transaction is completed and committed. You then have the choice to receive a receipt on paper, partly with the same information supplemented with extras that may be useful for you for administration and declaration purposes. It may also be useful when there are problems with the delivery. As soon as a new customer arrives and starts refuelling his car, all the displays are set to zero. The case has been closed.

This activity has two perspectives. The customer and the supplier who owns the pump. Both partly use the same information, only the usage context – the primary context – is different. You need the information for refuelling and, in the case you keep track of expenses, you can reuse the information for that purpose. The primary context of the supplier will also be the refuelling itself and the final payment transaction. In the back office of the supplier the information is needed for stock management and other forms of administration. When it is needed for business intelligence it can even become part of a data warehouse. The exchange of information between both parties takes place through the information on the display of the pump and the printed receipt. The information you receive on the receipt is also part of the information the supplier needs.

A short conclusion: we have an activity, i.e. refuelling, with a primary context triggered by information (signal about the empty tank) and with the necessary information objects that are exchanged via visual (display) and physical (print) ways. These are the three basic components of the aforementioned "something": a context, an activity and information objects, the main components of the information model that will be discussed in the next paragraph.

Leeuwarder Information model (LIM)

Interwoven trinity

The model based on the concept of "something" represents meaningful information as the only constant factor in a continuum of time and space

Just three entities are the essence of this model. First, the information has to be provided with context. To process information, it should be contained in structured form in an information object. For information to come into existence an event or transaction must take place. Part of an event or transaction are the different ways of exchange and communication.
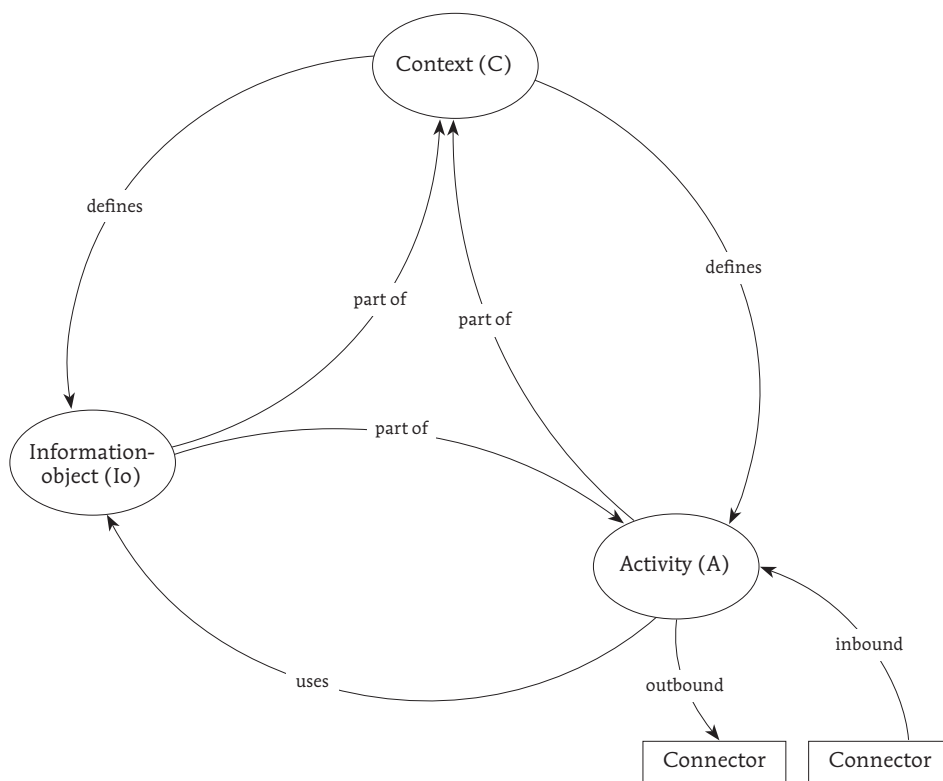
Figure 6. The Leeuwarder Information Model

The Information is central and has several invariant constant elements which underpin certain qualities. Qualities like meaningful based on context, structured as in the form of an information object and trustworthy by the way the context is described, the information object is structured and fixated, and the activity is executed. All qualities that are also essential composing parts of an archival record. In this model trustworthiness and veracity are derived qualities.

The context defines the setting in which the information has come into existence and at the same time, the information is used.

When speaking about use, I am talking about the primary use and therefore it is a primary context. Information can always be reused in other contexts. Within the context the characteristics of the information object and the activity are described and recorded.

The information object is just a container with the required information in structured form. The activity or act is a dynamic element that at the functional level processes and exchanges information within its context by using information objects. This involves exchanging information via one or more inbound (receive) and outbound connectors (send) with other activities. These other activities are again part of their particular context and information objects.

These three elements must be considered as an indivisible unit like a trinity. Together they make up meaningful information (MI). If one or two of the three elements are missing, there can be no meaningful information. If that occurs, the information is meaningless and therefore no more than useless data. These data may, of course, become meaningful by adding the missing elements artificially. Consequently, a new version is created. This new version constitutes a view of the original situation but is a reconstruction, and the information can therefore never be interpreted the same way as the original.

By means of following up the connectors that are part of the activity it is possible to make up a chain of meaningful information. This chain is an independent set of meaningful information. It can be compared to chemical elements that are bound together and make a chain to become a new chemical element.

The context

Context is about coherence in which notions, elements or concepts are intertwined, interwoven or connected. From its Latin origin, the verb *cotexo*, comes the meaning of interweaving and connecting. In the processing of information, context concerns the setting in which information is generated and used. Context is a concept that is at the essence of an archivist's discipline and therefore an important part of his field of study. Context is the essential element in understanding information and in the existence of information. Parts of that context are the objects in which information is included and the activities in which the information and information objects are processed.

There are many types of contexts to be considered. At first, a distinction can be made between primary and secondary context. Primary context is the context in which the information is originated or received. Context is secondary when already existing information is reused, so it is a use outside of the primary context of origin. Secondarily, context implies a point of view or perspective. A viewpoint determines in which way or with what purpose information is looked at. Also, it is the way the information is interpreted and appraised. There are infinite possibilities in this regard with for instance legal, spatial, economic, biological and theological and other professional views. All perspectives that can be intertwined with organisational and processual contexts.
It should be taken into consideration that use and reuse of information is never neutral or impartial. Even though it is often denied, there is always a subjective element present, namely the personal context, the insights of the person using the information. These insights are the results of upbringing, education, social position, and someone's own opinions and beliefs.

Moreover, it should not be forgotten that all these types of contexts or perspectives can also be found in the way in which devices and algorithms process information. Because they are the products of thoughts, ideas and opinions that originate from the various contexts.

In the model described above, the word context is used to describe or even define the connection between information objects and activities in a somewhat formal sense in order to understand the desired interrelationship. It is about the primary context.

This primary context consists of nine basic mandatory elements. The relation between these elements can be described in a semantic way. An activity only starts with a warrant or trigger. This activity will be an assignment executed by one or more actors (both humans and machines). The actors work in the environment or domain of a process, project or other kind of activity. Each activity has a result or product. It is possible that there is no result, but then the result is the 'non-result'. Information is processed by using functionalities, this information is recorded in information objects. Techniques deliver many functionalities to use, manage and preserve information objects. The context defines and describes both the activity and the information object.

In contrast with some archival standards, in this model an actor is not a separate entity with its own characteristics, but it is deliberately part of the context. The reason is simple. An actor, be it human or non-human, is just a passer-by in the in a continuum of time where information is at the centre. Only moment and place determine its role as participant or spectator.

Element	Meaning
1) Warrant	The reason or assignment why information is processed.
2) Assignment	The elaboration of the warrant in terms of an activity or transaction.
3) Actor	Who are involved with the assignment and the information processing. This can be both humans and machines.
4) Domain	The kind of work context in which the assignment is executed.
5) Result	What kind of result, products or conclusions are to be expected or not.
6) Functionality	What is allowed, what is possible for the different types of actors.
7) Data/Information	What type of information is processed, Personal, subject, object, financial, geographic, juridical etc.
8) Information object	How is the information captured, structured and shaped.
9) Technology	What technical tools and instruments are available or needed to process and preserve the information and information objects.

Table 1. Context - Elements

The reason or warrant for the activity refuelling was the warning of the gauge. Refuelling the car became the assignment. As a driver I was one of the actors.
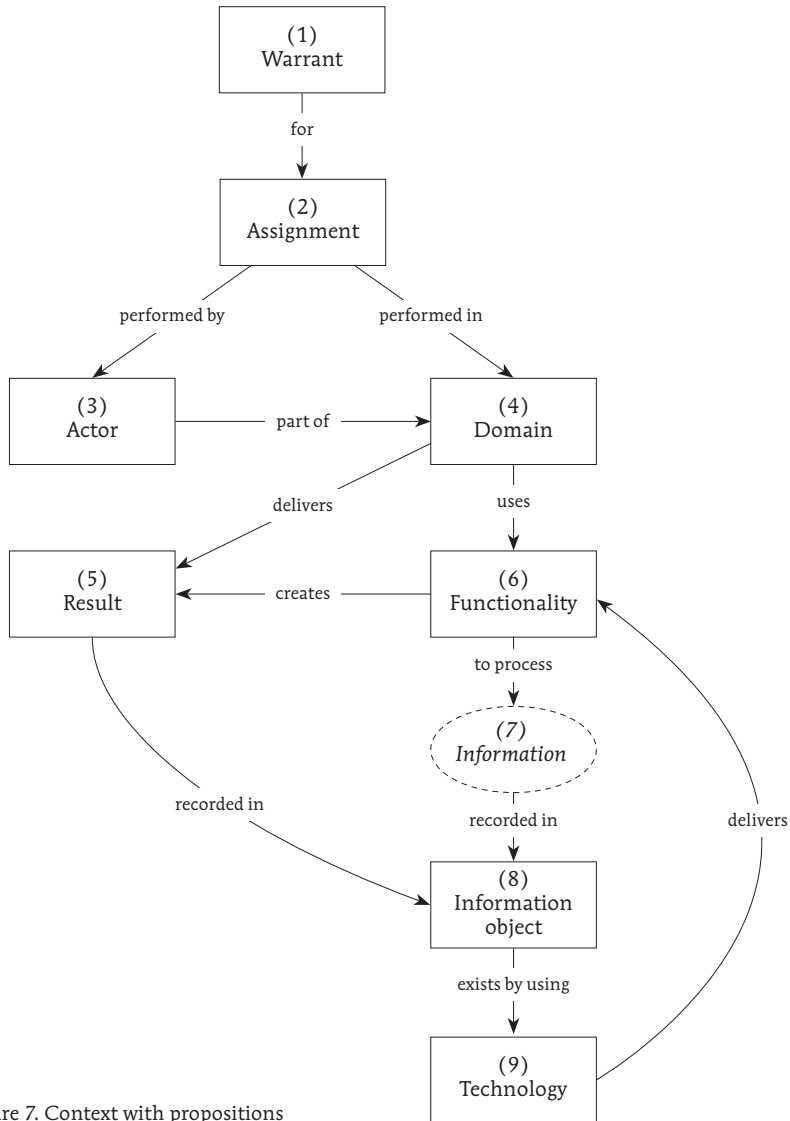
Figure 7. Context with propositions

Because of all the steps; parking, preparing the financial transaction, filling the tank and paying and getting a receipt, one could say this is a process. The other actors, the supplier, the oil company and the bank were not present in person, but they communicated through the ticket machine, the pump, and the display of the pump. The system of the gas station included the functionality to perform a financial transaction, to supply the necessary process information and of course the fuel. Display and receipt are examples of information objects. The pump and the ticket machine where both part of technology. Not visible, but certainly present was the technology of sending and receiving messages between the pump, the ticket machine and the systems of the supplier and the bank. As a client, I believed the whole transaction was trustworthy.

The information objects

Document

This part describes in what way information is recorded, structured and (sometimes) fixated. As a matter of fact, the word document can be used as a synonym for the concept information object. According to InterPares, a document is an indivisible unit of information constituted by a message affixed to a medium (recorded) in a stable syntactic manner. A document has a fixed form and stable content (Duranti & Preston, 2008). A document can be classified on the level of document types defined by its use or function e.g. evidential, informational, transactional.

A user experiences a document in the manner that this document is presented. It is a simple and unobtrusive object for the user. He sees it on his screen, he takes it out of the printer and holds the paper in his hands. What the user will probably not realise is that each document, both paper and digital, exists as an object of different ways of materialisation, techniques and dimensions or layers. For the user, a text and/or image in a document and the data carrier seem to be an inseparable unit, but they are not. What he does not realise is that a display on a fuel pump (or other devices) is in a broad sense also an information object i.e. document. In a sense he must be aware that information and carrier are detached. Knowing this he can focus on what is important: the information. The carriers are just the means to convey the messages.

Element	Meaning
1) Content	The message/information to be conveyed by the object.
2) Structure	The form of the message/information.
3) Context	The setting in which the object has a function and which is implicitly and / or explicitly present in the information object only. This context is therefore the result of the content and structure, and partly also the behaviour and technique.
4) Behaviour	Functionalities that must be executed to convey the message/information.
5) Technology	The tools needed to sustain the functionality and the information object and also to maintain and preserve the object. It consists of three main layers.

Table 2. Information object - Elements

A paper document can be copied to a digital image (scan), which still appears to be the same on the screen but is of a completely different order technically. For the user the text remains the same. The digitised text can be found via text recognition (OCR) and made digital as text and therefore usable for different operations. For example, by copying the recognised text to other documents.

This is possible because in fact, each document is composed of five characteristic elements. These elements are, (1) content, (2) structure, (3) context, (4) behaviour and (5) technology.

These five main elements are meant to hold information and convey a message. The intention of a message always depends on the context that is used within the object and the typology of the document. This typology itself is determined by the

combination of structure (2), content (1) and context (3). For example, the labels (metadata) on the pump at the gas station clarify the meaning of the data on the display.

Technical layers

An Information object consists of at least three technical layers. These layers are for (1) presentation, (2) transport and processing, and (3) storing and fixation. They allow recording information in documents and for those documents to be created, saved, edited, viewed and consulted, copied, sent, received and erased. At the same time, these layers are the determining factor for the behaviour of a document. These layers directly affect the other features of a document.

The presentation layer will display content, structure and context. This layer also ensures the way the content of a document must behave.
The transport layer provides for the processing of bit streams in such a way that the presentation is possible.
The storage layer ensures capture and maintenance of bit streams so that processing and presentation are possible.

Layer	Meaning
5.1) Presentation	Presenting the object to either a machine and/or a human being including the performance of behaviour.
5.2) Transport/processing	Transforming the factual technical form (e.g. bit stream) to make the message (content) presentable and therefore usable. E.g. Print to paper, show on digital interfaces like display screens, digital messages in XML between machines.
5.3) Store/Fixation	With analogue media like paper: the ink/toner attached to (or in) the medium (paper). Digital data: the bit streams, the zeros and ones on a physical data carrier or on more than one datacarrier which together make up a document.

Table 3. Information object - technological layers

As a practical example, this document is described below, using the above features and layers. This document is displayed on a screen of a PC or tablet or is available as a print on paper. This is possible because in the second layer, conversion of the digital source to the presentation layer (and vice versa) takes place. The digital source is visible as a file on the physical data carrier, but actually presents itself as a bit stream, a row of zeros and ones. Deep inside the technique of storage, the bit stream that makes up the document can be found, but unusable and unrecognisable without the support of software to operate the hardware and to make the document readable.

Within a word processor such as MS Word, content, structure and behaviour can be edited, modified, and deleted. Saving this document from MS Word to a PDF/A file, the content and structure are maintained. Only a small portion of the behaviour will be available, such as using links and going to a paragraph or a note with one click. In the paper copy this behaviour is not available, the reader can only browse more easily using his hands and eyes and he can write down his annotations.

The layers that make up the content cannot be seen separately from the technique. After all, the technique is instrumental in ensuring the use and survival of all components

Behaviour and compound documents

Behaviour, the fourth of the five characteristics of a document, exists by one or more embedded functionalities in a document. It is in a way part of the technology and is used to generate the presentation of the content and structure. Examples are animations or video and/or audio streams in presentations, generated charts, and execution of macros and formulas in spreadsheets. Another example is an embedded spreadsheet in a text document, which can be opened with just a click of the mouse. Also, hyperlinks – visible and embedded – to other documents or websites are part of the behaviour. The examples are uncountable. It is the element that makes a document dynamic.

The situation becomes more complex when there are compound documents that present themselves as one, but are made up of separate components. These components are often different digital documents on their own. In many cases, these individual components are made up of different file formats. A lot of compound documents are already around us, although not everyone recognises them (Ford, 2015). Examples of compound documents are email messages with attachments, digital documents with digital attachments, web pages, digital documents with links to other documents, and games.

Dynamic databases with queries and algorithms that are also documents – and that in a sense can be characterised as register – must be placed under the category of compound documents that contain a lot of complex behaviour.

From a user's perspective there will not directly be a notion of a compound document. At the level of presentation, the user sees the document as a unit. This presentation depends on the three technical layers. Through the transport/processing route, the components are retrieved from the storage/fixation (the disks) by one or more applications and digitally forged to a temporally unit which is delivered to the presentation layer. These operations can be considered as behaviour (4).

A document is also a compound when links refer to other documents that are stored on internal or external servers as if they were attachments. At an abstract level – although difficult to understand – the total of these files constitutes one document. Examples are web pages.

These types of documents are a major preservation challenge for archivists because the management of these information objects is mostly beyond their reach. The complexity lays in the requirement that the presented unit, the total of the components of a compound document, must be preserved and maintained until the date of its disposition. An issue that may arise are broken links. Not all links are permanent: they can be modified, the external files are missing or have disappeared, or other versions of the external files are provided with the same link. Reference rot with content drift and link rot are looming and menacing perspectives.

The activity

Time and steps

Within an activity, as described or defined in the context, the information objects are used and exchanged with other activities. Activities that are part of their own set of meaningful information (MI). Please note that when the word activity is mentioned, it is not about the (trans)action itself but about what happens with information and information objects during that activity.

Element	Aspect
Time	Start Moment(s) End
Exchange/Communicate	Send (message, make available, transfer) Receive (message, fetch, acquire)
Steps	Assess Capture Process Keep/Retain/Dispose

Table 4. Activity - Elements and aspects

From both the previously mentioned concept of "something" and the example of the activity of refuelling, the core elements in question are perhaps self-explanatory and visible. These are the time factor, the sequence of the steps within the information is processed and the moment the exchange of information takes place. The aspects of time are the beginning, the end and the moments in between.

There are four aspects in the processing of information that are most likely default. Firstly, it is the step of assessing and appraising to get an answer to the question if the information in the information object has to be used and therefore ingested. When the answer is positive, the information object can be ingested and registered. Then the information can be interpreted and used within the context of the activity. Also, information can be changed, added or new information objects can be created. Within the fourth default step the used or created information is saved and kept for reasons of availability and accessibility or is perhaps disposed, transferred or destroyed.

The steps when information is assessed, appraised, used and eventually saved or destroyed are usually not sequential. An often-enforced sequence can be found especially in formal situations where evidence, accountability, control and supervision play an important role. In other cases, it is therefore not strange when the steps mostly seem to run across. For example, during the time of use, the used information can directly be destroyed.

The exchange element

It is important to understand what happens when information is exchanged on a certain level of abstraction. Certainly, for archivists with that understanding and their deep rooted archival knowledge it is possible to determine which parts of digital information must be archived. It is also possible to ascertain conditions and requirements the exchange should meet. For an archivist the message is always an information object, be it a document, a set of documents or datasets. Therefore, this information object has the same characteristics as every other information object; there is a content, a structure, a context, a behaviour and technology.

Messages and additional metadata for sending and receiving are wrapped in a container, an envelope which is in again an information object.

The exchange model is a straightforward broad concept; there is always a message, a sender, a receiver, and a channel. It is always a form of communication between one or more parties (actors and processes). It is necessary that information will get from place a to b unhindered and undamaged. In many cases, the received message after unwrapping and accepting becomes part of a new context. Although this is of concern for the sender, it is nevertheless beyond his reach. It is just a fact. As a result, the intention of the transmitter may not always occur to the recipient. Therefore, transmitted information must be provided with sufficient metadata to convey the intention.

From the perspective of an archivist, the communication model may look as in the scheme below
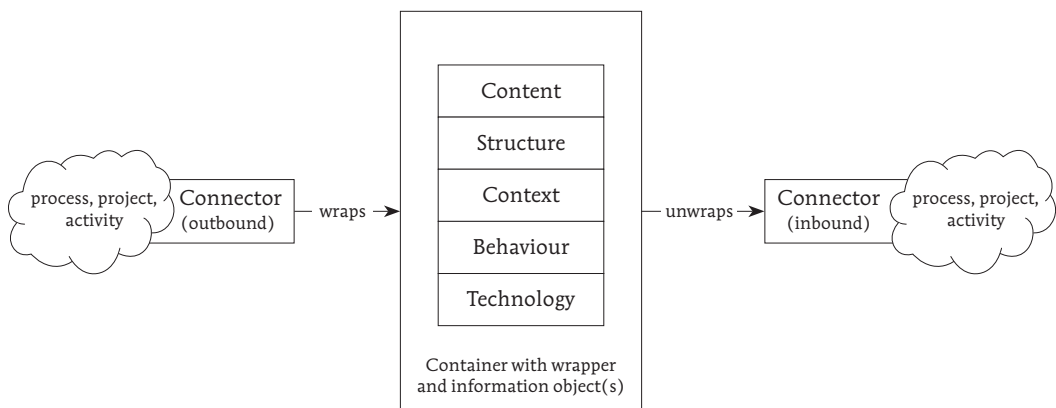
Figure 8. Exchange model

This kind of exchange takes place at the moment when, for example
Information is exchanged directly between or inside processes, projects or other activities, high standards of interoperability are necessary;
Information is exchanged via messaging, chat, correspondence via surface mail;
Information is published on websites; open data sets are made available;
Transferring information from one actor to another, an action that involves a move or migration, where the original is deleted (destroyed) after a successful transfer. High standards of interoperability are necessary.

Several communication models exist; they are models that indicate how information and data are transmitted. On the technical level, the best-known model is from Shannon (Shannon 1948). This model is about how information can be obtained from the sender by the receiver. The transmitter puts the message in a signal to be transmitted (encode), after transmission the receiver extracts the message (decode). The context of this model is the telecommunications.

Another communication model comes from linguistics. This model can be found in the work of the linguist Roman Jacobson about his theory of communicative functions of language (Jacobson, 1960). In this model a channel encodes a message in a code. Then the message is intentionally send via a medium to a recipient who decodes the message. The message must have a context that can be referred to and that makes it understandable to the recipient. This linguistic model indicates what is needed to get the contents of a message from sender to recipient while maintaining interpretability.

For all models, it is important that a message is transmitted without loss – be it in a technical, linguistic or archival sense. The nice thing is that for all three the core is the same, for example, there is always a transmitter and receiver, and in all three cases a message is converted (packed, encoded, encrypted, wrapped, unpacked, decoded, decrypted, and unwrapped). In the core all three use channels. The difference is the origin of the models, each with its own characteristics.

The time difference between sending and receiving is invaluable. There is always a time difference between the time of sending and the moment of receiving information. In many cases, that time difference will be minimal and almost negligible because of the state of the art of the technique. But there may be cases where it is necessary to know if someone could have been aware of a specific situation at some point. In some transactions, like in stock markets, the quote 'time is money' has a literal meaning.

Semantic information

As said before, in these turbulent times of digital change an archivist needs a sustainable fundamental layer on which he can build instruments to do his work. Parts of that foundation may be found in the philosophy of information. For this it is necessary to find the conceptual relationship between archival science and information philosophy.

According to Floridi, philosophy of information is the philosophical field concerned with the critical investigation of the conceptual nature and basic principles of information, including its dynamics, utilisation and sciences (Floridi, 2009).

For a possible connection with the philosophy of information, the Leeuwarder information model (LIM) about meaningful information as described above will be used. From this starting point it is a small step to a model explained by Floridi (Floridi, 2010) where he introduces semantic content with elements as meaningful and well formed. According to Floridi is information semantic when it is also truthful and therefore true.

Roughly speaking there are connections between two elements that make up semantic content from philosophical perspective and meaningful information from the archival perspective. These are from the viewpoint of the philosopher respectively "meaningful" and "well-formed" and from the viewpoint of the archivist "context" and "information object". Because the context delivers meaning to information, both "meaningful" and "' context" are on the same level. An information object contains information in such a structured form that it can be called well-formed.

On the point of semantic information there is a more principle difference between the philosopher and the archivist. An archivist can only ensure and ascertain that an information object under his control has the same quality as when it was ingested, it is trustworthy with guarantees about the integrity. An archivist cannot and from an ethical viewpoint may not deliver a statement about truthfulness of the information.[2] The answer to the question if the information can be assumed truthful can only be given by the user of that information acting from his various contexts. Anyway, knowing this difference, one could say that under certain conditions the archival trustworthiness is at the same level as the philosophical truthfulness. This means that a real connection between both fields of study is just as well possible.

Model Philosophy of Information (PI)	Leeuwarder information model (LIM)
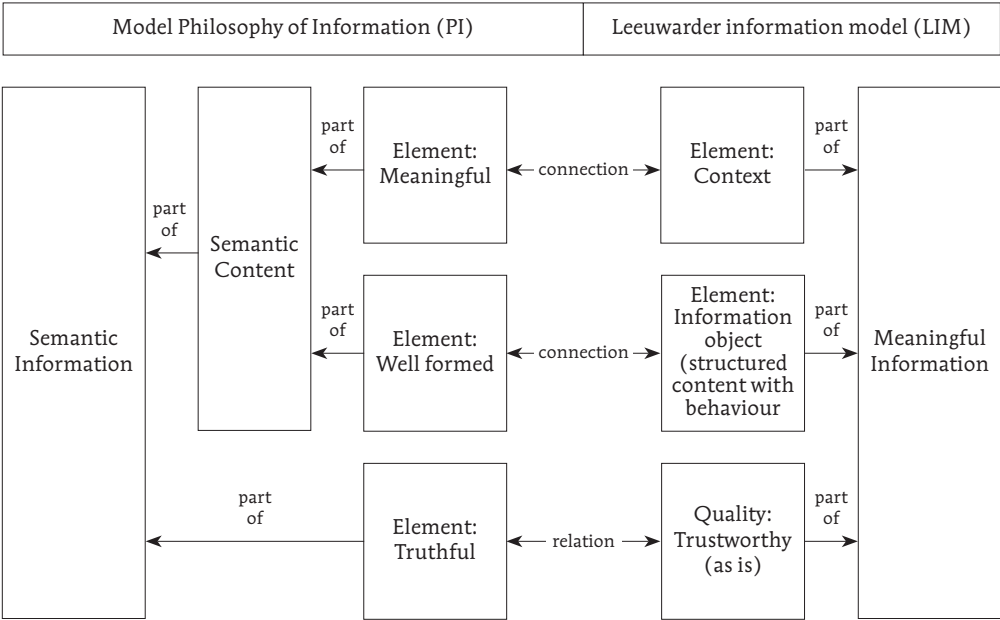
Figure 9. Bridging the gap

---

2 This statement does not release an archivist from a moral obligation to help a user as much as possible with his research in the archives.

This connection can be used as a bridge that spans the virtual gap between both sciences. This can be a route for archivists to enter the realm of the information philosopher. A world of very fundamental research on information and concepts of information with its own semantics which may be useful for the archivist who should be aware of the consequences of the infosphere on his discipline. At the same time for the philosopher the knowledge of the archivist as appraiser, broker and curator of meaningful or semantic information becomes available. As shown above there are certainly connections between the concepts of the philosopher and the archivist. These connections need further research.

Literature

Adriaans, Pieter (2013). *Information* (lemma). In *The Stanford Encyclopedia of Philosophy*. Retrieved from https://plato.stanford.edu/archives/fall2013/entries/information/

Acker, Amelia (2016). When is a record; A Research Framework for Locating Electronic Records in Infrastructure. In A. Gilliland, S. McKemmish, A.J. Lau (eds.) *Research in the Archival Multiverse* (page 288-233). Clayton, VIC: Monash University Publishing. http://dx.doi.org/10.26530/OAPEN_628143

Allo, Patrick, Bert Baumgaertner, & The ∏ Research Network (2013). *The Philosophy of Information; An Introduction*. Retrieved from http://socphilinfo.org/news/teaching/412-philosophy-information-introduction-new-extended-version

Beavers, Anthony F. (2012). *A Brief Introduction to the Philosophy of Information*. Evansville, IND: University of Evansville, Philosophy / Cognitive Science. Retrieved from http://www.academia.edu/1230161/A_Brief_Introduction_to_the_Philosophy_of_Information

Duranti, Luciana (1998). *Diplomatics; New uses for an old science*. Lanham / London: Society of American Archivists and Association of Canadian Archivists in Association with The Scarecrow Press Inc.

Duranti, Luciana, & Randy Preston (2008). *International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records*. Padova: Associazione Nazionale Archivistica Italiana. Retrieved from http://www.interpares.org/ip2/book.cfm

Floridi, L. (2009). The Information Society and Its Philosophy: Introduction to the Special Issue on "The Philosophy of Information, Its Nature, and Future Developments". *The Information Society, 25*(3), 153–158. https://doi.org/10.1080/01972240902848583

Floridi, L. (2010). *Information: A Very Short Introduction*. Oxford: Oxford University Press.

Floridi, L. (2013). *The Ethics of Information*. Oxford: Oxford University Press.

Floridi, L. (2014). *The Fourth Revolution; How the Infosphere is Reshaping Human Reality*. Oxford: Oxford University Press.

Floridi, L. (ed.) (2015). *The onlife manifesto; Being Human in a Hyperconnected Era*. Heidelberg: Springer. Retrieved from https://link.springer.com/book/10.1007/978-3-319-04093-6

Ford, Paul J.(2015). What is code? *Business week*, 11-jun-2015. Retrieved from https://www.bloomberg.com/graphics/2015-paul-ford-what-is-code

Gilliland, A., S. McKemmish, & A.J. Lau (2016). *Research in the Archival Multiverse*. Clayton, VIC: Monash University Publishing. http://dx.doi.org/10.26530/OAPEN_628143

Gleick, James (2011). *The Information; A History, a Theory, a Flood*. New York: Knopf Doubleday Publishing Group.

Glaudemans, Arnoud, Rienk Jonker, & Frans Smit (2015). The right people and the right forces need to get together,: Interview with Prof. Luciano Floridi. *Archievenblad*, Januari 2015. Retrieved from http://www.youblisher .com/p/1048348-Archievenblad-uitgave-van-de-Koninklijke-Vereniging-van-Archivarissen-in-Nederland/

Heinlein, Robert (1961). *Stranger in a strange land*. New York: Puntnam's Sons.

Jacobson, Ramon (1960). *Closing Statements, Linguistics and Poetics*. In Thomas A. Sebeok (ed.), Style In Language (pp. 350–377). Cambridge MA: MIT Press.

Maes, Rik (1999). *A generic framework for information management (PrimaVera Working Paper 99-03)*. Amsterdam: Department of Accountancy & Information Management. Retrieved from https://www.researchgate.net/publication/242321998_A_Generic_Framework_for_Information_Management

McKemmish, S. (2005). *Archives: Recordkeeping in society*. Wagga Wagga: Centre for Information Studies, Charles Sturt University.

Muller, Samuel, Johan A. Feith, & Robert Fruin (2003). *Manual for the arrangement and description of archives* (Transl. of the second ed. by Arthur H. Leavitt, with new introductions by Peter Horsman, Eric Ketelaar, Theo Thomassen, and Marjorie Rabe Barritt). Chicago: Society of American Archivists. Retrieved from https://hdl.handle.net/2027/mdp.39015057022447

Shannon, Claude (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal, 27*, 379-423 and 623-656.

Some standards

ISO 23081-1:2006, Information and documentation – Records management processes – Metadata for records – Part 1: Principles

MoReq2010 - Modular Requirements for Records Systems, Volume 1 - Core Services & Plug-in Modules, DLM Forum Foundation, Brussel, 2011. Version 1.1

GEOFFREY YEO

Information, Records, and the Philosophy of Speech Acts

Introduction

Information is a prominent theme in 21st-century archival discourse. In English-speaking countries, records professionals often affirm that distinctions between information and records are increasingly blurred (Clarke, 2009, p. 121; Trombley, 2016, p. 50); many records managers, and some archivists, now use the words 'records' and 'information' almost interchangeably. According to British records manager Steve Bailey (2008, p.59), 'any student of archives or records management will tell you' that 'all records are information'. When the international standard for records management (ISO15489-1:2016) defines a record as 'information created, received and maintained … by an organization or person, in pursuance of legal obligations or in the transaction of business', it suggests that records are a kind, or perhaps a subset, of information.

Another perspective sees information as the *content* of a record. For American writer William Saffady (2011, p. 236), a record is 'an information-bearing object'; for Frank Boles and Mark Greene (2001, p. 435), 'information is embedded in a record'. Seen in this light, information is a commodity that can be captured and stored within records, from which it can be retrieved when it is needed. Many practitioners accept the view, commonly promoted in records management literature, that a rigorous record-keeping programme can ensure that the contents of records are full and accurate (Financial Conduct Authority, ca.2015, p. 1; Pember & Cowan, 2009, pp. 5-6; State Records New South Wales, 2004). Advocates of this view often associate information with notions of *fact*. They affirm, or assume, that accurate and complete factual information can be objectively captured in records, and that – provided the records are well managed – the information held or set down in them will then be transmitted unambiguously to users.

In the literature of the wider information field, information is sometimes described as 'a collection of facts' (Stair, Moisiadis, Genrich, & Reynolds, 2011, p. 6). According to one textbook, information 'is not subjective or biased but … factual and impartial' (McGonigle & Mastrian, 2012, p. 20). Comments such as these presuppose a world in which information is context-independent and language can be made to correspond to external realities. As Ciaran Trace (2016, p. 56) has noted, the seemingly factual nature of information obtainable from organisational records is often taken for granted by business-unit employees, as well as by records managers and archivists. However, growing numbers of scholars have argued that information is always a construction, and does not exist independently of the situations in

which it is produced (Cornelius, 2014, p. 190; Gitelman, 2013, pp. 2-3). From this perspective, claims that information can be objectively accurate and impartial are misguided attempts to conceal its contingent role. Arguments of this kind call into question many of the ideas voiced in records management literature, about information, accuracy, and the relevance of these notions to an understanding of records and their contents.

In addressing these issues, this essay explores connections between records and information, and examines how a number of other theoretical perspectives (notably the philosophy of speech acts) can enhance our comprehension of records and their functioning. It considers how notions of 'information' might relate to a view of record-making and record-keeping that takes account of speech act philosophy, and it concludes that records have social as much as informational roles.

Distortion, error, and conscious production

Medical records offer a useful starting-point. Despite the recognised importance of careful record-keeping in the medical sphere, studies of the content of medical records have revealed numerous inadvertent mistakes made by clinicians hastily entering details of diagnosis or treatment (Bowman, 2013; Lloyd & Rissing, 1985). But many distortions have been found to occur for other reasons. When clinicians record what their patients have told them, cultural or organisational pressures often come into play and lead them to condense the patients' words into formal professional narratives. Although records ostensibly report what patients said, the patients' vocabulary is converted into that of the clinician, and 'what little of patients' voices appears in those reports appears only on the health professionals' authority' (Poirier, 1999, p. 34; cf. Berg, 1996, pp. 505-507). The potential for accidental error, ambiguity, or deliberate misrepresentation can never be excluded.

Similar constraints are apparent outside the medical world. Investigations of social work case files have shown that they frequently document tasks, goals, and their accomplishment, but cannot capture the everyday complexity of the personal lives that prescribe and determine them (Floersch, 2000, p. 179). Other researchers have reported that files created by parole officers cannot fully reflect the behaviour of parolees and often give undue emphasis to incidents that support officers' desired outcomes (McCleary, 1977). In many organisations, minutes of meetings tend to capture only formal aspects of a meeting and omit controversial or politically sensitive aspects (Whittaker, Laban, & Tucker, 2006, pp. 104-106; Winsor, 1999, pp. 218-219). Some commentators affirm that records such as these are created by those in positions of power and remain largely silent about matters their creators prefer to overlook (Thomas, Fowler, & Johnson, 2017); others argue that lived experience can never be adequately reduced to mere textuality (Floersch, 2000, pp. 169-170; Lynch, 1993, p. 287).

Many studies have indicated that record-making practices are often directed at minimising any trace of digressions from formal procedures or showing the creators of records in the best possible light. Workers compiling timesheet records may seek to demonstrate compliance with official accounting rules rather than recording the precise hours they actually worked (Brown, 2001). An ambassador's dispatches

may be worded to give an exaggerated sense of achievement or to conform with the expectations of the government at home (Tosh, 2015, p. 107). Even if we set aside the most blatant examples of deceit, such as the falsification of police records discussed by John Van Maanen and Brian Pentland (1994, pp. 73-75), it is clear that the self-interest of record creators and their desire to operate successfully within bureaucratic cultures often leads them to construct records in ways that are intended to create particular impressions or advance particular points of view.

Much of the early work that points in this direction was undertaken in social science disciplines. As long ago as 1980, social scientists Nancy Cochran, Andrew Gordon, and Merton Krause concluded that, although 'the commonly accepted understanding ... is that records preserve information about the incidence of ... events', records are 'proactive rather than ... simply descriptive' and are preceded and shaped by the plans, goals, intentions, and assumptions of their creators (1980, pp. 5-6). Archival scholars did not generally come to share this view until rather later (and it is a view still rarely acknowledged in the practice-focused literature of records management).

Awareness that records 'are not neutral, factual, ... [but] are designed ... to produce an effect in some kind of audience' (Van Maanen & Pentland, 1994, p. 53) reached archival scholarship with the postmodern 'turn' in the archival literature of the 1990s. Terry Cook was among the first archivists to acknowledge that all records are conscious products. When we encounter records or archives, he said, we can be beguiled into assuming that they convey neutral data or information (Cook, 1994, p. 319), but behind the record lies the record-maker and the contexts of activity in which the record was produced, and 'archivists want to know ... not just what was communicated, but when, by whom, to whom, where, how, [and] why' (1994, pp. 302, 312). At the start of the new millennium, Cook (2001, p. 25) linked these ideas to a postmodern view of records, when he characterised archival postmodernism 'as focussing on the context behind the content; on the power relationships that shape the documentary heritage; and on ... business-process conventions as being more important than ... informational content'.

Diplomatic scholarship

Of course, it is not only archivists sympathetic to postmodernism who emphasise process and activity. Instead of simply associating records with information, writers engaging with archival science have often stressed that records 'are inextricably connected with activity' (Eastwood, 2017, p. 16), although they have not always agreed what that connection might be or how it might operate.

Many archivists, particularly in countries whose legal traditions are those of Roman-influenced civil law, have believed that explanations of the connections between records and activity can be discovered within the discipline of diplomatic scholarship. This discipline, built on the work of the 17th-century monk Jean Mabillon, is known as 'diplomatic' in Europe and as 'diplomatics' in North America. It seeks to define objective means of assessing the authenticity of documents that have legal consequences, particularly charters, diplomas, and other documents that attest to the granting of rights. Over time, legislative systems have

come to accept that individuals or institutions can grant rights – and impose obligations – through the issuance of an appropriate document, and practitioners of diplomatic give the label 'dispositive' to documents that substantiate actions of this kind; documents that supply evidence of actions that were 'complete before being manifested in writing' are labelled 'probative' (Duranti, 1998, pp. 65-66; Duranti & Preston, 2008, pp. 811, 830). Both labels are confined to documents whose written form is legally (or 'juridically') recognised.

Although the English word 'record' is not native to the civil-law traditions in which diplomatic scholarship has flourished (Yeo, 2015, p. 315), diplomatists working in an English-speaking environment have equated records with the documents that are the focus of diplomatic study, and have thus been able to claim that the field of diplomatic 'categorizes ... records according to their relationship with the acts that caused their creation' (Duranti, 2010, pp. 1594, 1596).

Many diplomatists have also sought to expand the scope of their field beyond the study of documents or records with legal consequences, to encompass records relating to any aspect of human affairs (Boyle, 1992, pp. 87-88; Duranti, 2010, p. 1594). Recognising that records now play a variety of roles outside the legal arena, they have attempted to apply or adapt diplomatic principles and methods to new forms of record in the contemporary world. In Germany, for example, 20th-century scholars set out to extend the range of diplomatic beyond *urkunden* (diplomas and charters) to *akten* (office files). In some quarters, however, this endeavour generated a critical response; because diplomatic criteria remained the norm and office files did not fulfil this norm, 'they were simply treated as the *other* of diplomas', as non-*urkunden* (Vismann, 2008, p. 75).

Much the same can be said of Luciana Duranti's proposal to cater for an expanded scope of diplomatic by supplementing the traditional categories of 'dispositive' and 'probative' with additional categories of 'narrative' and 'supporting' records. In Duranti's writings, 'narrative' records are described as constituting 'evidence of activities that are juridically irrelevant', and 'supporting' records are said to constitute evidence of activities that are juridically relevant but do not 'result in a juridical act' (Duranti, 1998, pp. 68-69; Duranti & Preston, 2008, pp. 825, 840).[1] Juridical relevance – or the presence or absence of legal consequences – appears to be the yardstick by which all records and activities are assessed. Although the questions that diplomatic scholarship asks (can I trust this record? do I think it is authentic?) are pertinent to records of any kind, its analytic approach to answering them arguably works best when applied to the legalistic records for which it was first designed. For diplomatists, such records remain the prototype,[2] and other forms of record are judged by what they lack in relation to them.

Speech act theory

In earlier work (Yeo, 2010), I argued that concepts of 'speech acts', developed in the second half of the 20th century by the Oxford-educated philosophers J. L. Austin and John Searle, offer a fruitful alternative approach to understanding the connections between records and activity. Speech act theory affirms that, in speaking or writing under the right conditions, we can perform certain kinds of act.

Although it was discussed occasionally in archival literature before 2010 (Brothman, 2002, p. 320; Henttonen, 2007; Underwood, 2008), the applicability of speech act theory to archival science has not been widely addressed. The studies by Austin (1962) and Searle (1969; 1979; 1995) tended to give more emphasis to speaking than to writing, and Searle's promotion of the label '*speech* acts' has probably obscured the relevance of their concepts to a discipline such as archival science, which is largely concerned with written documents. Nevertheless, speech act theory has been explored and adapted by numerous scholars in philosophy and other fields, including many who have applied it to acts performed by means of written texts (Cooren, 2004; Doty & Hiltunen, 2009; Ferraris, 2013; Kurzon, 1986; Lee, 1988; Smith, 2014; Winograd & Flores, 1986).[3]

In his seminal work *How to do things with words* (1962), Austin challenged the assumption that the sole function of language is to transmit information. His examples of what he called 'performative' uses of language included utterances such as 'I pronounce you guilty', 'I bequeath you my watch', 'I apologise', and 'I name this ship'. He noted that a ruler or official who says 'I name this ship the *Queen Elizabeth*' does not simply tell her listeners what the ship is to be called; by uttering these words, she performs the act of giving the ship its name. Likewise, when I write 'I apologise' in an email, I do not merely send information about an apology; I perform the act of apologising.

Broadly similar ideas had been proposed earlier in the 20th century by the American pragmatic philosopher George Herbert Mead, who wrote that 'language does not simply symbolize a situation or object which is already there in advance; … it is a part of the mechanism whereby that situation or object is created' (1934, p. 78). At a yet earlier date, the German philosopher Adolf Reinach (1913), and before him the Scotsman Thomas Reid (1788), had attempted to analyse language in terms of social acts. Austin (1962) likewise noted that to ask a question, give an order, or make a promise is to perform an act, and he proceeded to argue that stating a proposition is also performing an act. Making a statement is no less a social act than apologising, bequeathing a watch, or naming a ship.

After Austin's death, his thinking was further developed by his former pupil Searle, who propounded a taxonomy of speech acts that many subsequent writers have found useful. Searle (1979, pp.12-20) identified five basic categories of speech act: *assertives, expressives, directives, commissives*, and *declaratives*. In an *assertive* act, speakers or writers state a proposition about how things are, were, or will be;[4] in an *expressive*, they express their feelings or attitudes; in a *directive*, they ask a question or attempt to get someone to do something; in a *commissive*, they commit

---

[1] These distinctions may be more intuitive to archivists working in civil-law jurisdictions than to archivists in common-law countries unfamiliar with notions of 'juridical acts' and 'juridical relevance'. However, the categorisation of 'narrative' and 'supporting' records that Duranti invented does not seem to have been widely adopted outside her own research and the work of the researchers she has directed. More recently, she has identified 'instructive' and 'enabling' as further categories that are said to be characteristic of digital environments (Duranti & Preston, 2008, pp. 814, 819; Duranti, 2010, p. 1596).

[2] For an account of prototype theory and its application to understandings of records, see Yeo (2008).

[3] Scholars such as Zsolt Bátori (2015) have also considered the application of speech act theory to images and visual resources, but these lie beyond the scope of this chapter.

[4] In Searle's approach to speech act theory, other types of speech act besides assertives are assumed to have 'propositional content' (Searle, 1979, pp. 14-20; cf. Hanks, 2015, pp. 200-204), but in this chapter the term 'proposition' is used in its general sense of 'the content of an assertion'.

themselves to doing something; in a *declarative*,[5] they 'make changes in the world' (Searle, 1999, p. 150) through their utterance, as in the case of utterances that declare war, dismiss an employee, adjourn a meeting, or bequeath a watch. In each case, the act is performed through representation and communication. According to Searle (1983, pp. 166-167), 'we can perform an act … by representing ourselves as performing it', if we make the representation under the right conditions and communicate it to someone (or perhaps to a storage system that someone can use).[6] Like most attempts at categorisation, Searle's taxonomy has been subject to criticism (Hancher, 1979; Suchman, 1994), but it remains influential among philosophers and analysts of speech acts.

It is also possible to apply Searle's categories to written records. As a preliminary to further analysis, it can be suggested that:

- Assertive records make statements or assertions; they are representations of claims that humans, or computers programmed by humans, make about the world. Such claims may be about past actions ('I assert that I attended the conference') or past, present, or future situations ('I confirm that the company has 200 employees'). In practice, phrases such as 'I assert that' and 'I confirm that' are usually omitted. In creating an assertive record, a human may merely write 'I attended the conference'; a computer may merely generate the statement 'Number of employees = 200'.
- Expressive records (such as 'I apologise') do not make claims about the world; they 'take it for granted, and simply react to it' (Hancher, 1979, p. 3).
- Directive records range from those that merely represent the asking of a question ('Were all customers satisfied?') to those giving commands or orders ('Contact headquarters immediately!').
- Commissive records also vary in strength; a record creator may make weak or generalised commitments ('I will return next month'), utter promises ('I promise that I will return'), or enter into formal contracts enforceable at law.

A declarative record represents the making of 'changes in the world', because communication of the record creates a state of affairs that did not obtain previously. According to American philosopher Nick Fotion (2000, p. 51), declaratives 'bring about a change in status or condition just in virtue of being uttered successfully'. As Searle (1979, p. 17) observed, 'if I successfully perform the act of appointing you chairman, then you are chairman; … if I successfully perform the act of declaring a state of war, then war is on'. He could have added that if I successfully perform the act of bequeathing you my watch, then my watch is yours; the act creates a new right of ownership. Declaratives have close parallels to 'dispositive' records as understood by diplomatists.

However, it can be argued that declaratives are not the only records that operate in this way. The issuance of commissive records such as contracts and promissory notes makes 'changes in the world' by creating obligations that previously did not exist. Some directives also create obligations; summonses, for example, oblige their recipients to attend a tribunal. Searle (1995, p. 70) referred to the responsibilities and rights that are created in this way as 'deontic' phenomena, a term he derived from the work of Finnish philosopher Georg Henrik Von Wright (1951). They are

not limited to those recognised at law or requiring legal formalities; an obligation, for example, might arise simply from a promise, made by email, to attend a meeting or visit a sick relative. Obligations, ownerships, permissions, and other deontic entities are social realities; they exist in virtue of collective acceptance (Searle, 2008, p. 27), and they remain in existence only for as long as they continue to be accepted or recognised.

In the words of philosopher Barry Smith (2008, p. 44), deontic entities have 'an anchorage in the realm of records and representations'. In forming deontic entities, as Searle (1995, p. 85) noted, 'the characteristic devices used are … deeds, bills of sale, registration papers, wills,' and other types of official document. In Western societies, issuing such documents counts as conveying properties or creating rights. Rights can be expected to persist for a period of time; they outlast the moments of issuance of the documents that create and confer them. But the documents also endure long after their moments of issuance, and laws or societal conventions permit them to function as evidence of title; if I have the deed by which you sold me your house, I can use it to demonstrate that the rights to the house are mine.

This practice can sometimes be taken further, and documents can become negotiable instruments. In such instances, a transfer of a physical record constitutes a transfer of the rights that were created when the record was issued. For example, when a debtor has given a written promise to pay a sum of money to a creditor, a transfer of the promissory note to a third party will transfer the right to receive the sum in question (Jordan, Warren, & Walt, 2000). Far from being simply information about a debt or other obligation, such records seem to embody the obligation itself. Like the prior notion that a promise imposes an obligation on the promiser, the notion that obligations can be embodied in records is a human construct, but constructs of this kind help to shape lives, coordinate human behaviour, and support social and economic well-being.[7]

Of course, rights and obligations are not restricted to literate peoples, and in any human society they can – in principle – be generated orally, without creating written records. The earliest written records of property rights were assertions that a purchaser had acquired a property that had been conveyed to him by oral or physical means; they reported a 'change in the world' that had already occurred when the record was created. Over time, the idea grew up that such records, if created using exact written forms, would possess evidential weight in the event of legal disputes.

---

[5] Searle's use of 'declarative' (or 'declaration') in his taxonomy has sometimes caused difficulty. For scholars in linguistics, the connotations of 'declarative' are often closer to what Searle called an 'assertive'. Searle's usage also differs sharply from the use of the word 'declarative' in what is called a 'declarative act' or *acte déclaratif* in certain civil-law jurisdictions. However, Searle's usage is widely followed in the literature on speech act philosophy, and I have retained it here.

[6] In her critique of my earlier work, Anneli Sundqvist (2011, p. 283) appears to overlook this point, which is argued more fully in Yeo (2007, p. 338) and (2010, pp. 100-101). Representations are not merely what Sundqvist, following Marx Wartofsky, calls 'secondary artefacts', which describe or 'reflect' actions that have occurred at some earlier time; they can also help to constitute an action.

[7] Besides undermining the simple identification of records with 'information', the analysis given here negates the contention by Victoria Lemieux (2014, p. 77) that, because all records are cultural constructs, all are representations of 'beliefs'. Searle (1979, p. 4) observed that 'a man who states, explains, asserts or claims … expresses [a] belief', and this seems broadly correct; we may accept that creators of 'assertive' records express beliefs about how things were, are, or will be. However, 'directive', 'commissive', and 'declarative' records operate differently; they are underpinned by human conventions, but do not openly express or represent beliefs about the world.

As noted above, diplomatists attribute a special 'probative' status to a record that offers a formalised report of a completed transaction. From the perspective of speech act theory, however, such a record simply represents an assertive speech act, albeit one to which legal systems may accord particular recognition (Searle, 1995, p. 85).[8]

When we find weight attributed to a record that asserts a proposition, we may detect a paradox. On the one hand, scholars doubt whether propositions are ever objective; on the other hand, legal systems acknowledge that a record asserting that a specific transaction has occurred can provide litigants with strong evidence, which may help them prove their rights in court. It seems reasonable to conclude that, whatever our view of the objectivity of a proposition, we should not consider its assertion ineffectual. When we examine the making of 'changes in the world' by means of speech acts, we may affirm that 'every language act has consequences' (Winograd & Flores, 1986, p. 59); even an assertive act can achieve results that would have been impossible if the assertion had not been promulgated.

Speech acts and genres

Several scholars have drawn parallels between speech act philosophy and genre theory, which some archivists have recently heralded as offering a new theoretical framework for investigating records and their modes of creation (Foscarini, 2013, unpaginated; 2014, pp. 6, 23; Oliver, 2017, p. 95). According to American scholar Charles Bazerman (2012, p. 383), genres are 'typified forms of utterances'. While each individual speech act is unique, communicative activity depends on our ability to invoke stabilised *types* of utterance. In Bazerman's words, speech acts are carried out 'in patterned … textual forms or genres, which are related to other texts and genres that occur in related circumstances' (2004, p. 311). The recognition (in taxonomies such as Searle's) that speech acts can be categorised into types is paralleled in what Bazerman called the 'typification' of genres, such as business letters, reports, order forms, and contracts. Genre theorists have argued that each genre corresponds to a particular type of social activity (Gardiner, 1992, p. 81) and is used in response to specific recurrent situations (Yates & Orlikowski, 1992, p. 301).

Connections between genre theory and speech act theory are often associated with ideas about 'speech genres' put forward by Russian scholar Mikhail Bakhtin and his circle in the mid 20th century. Bakhtin's work, which remained unpublished for many years after he composed it, does not use the term 'speech acts', but it displays many of the same beliefs about language that characterise speech act theory. Like speech act philosophers, Bakhtin distinguished utterances from sentences (1986, p. 73) and insisted that 'the real unit of … communication' is the utterance (1986, p. 71). He argued that utterances cannot be regarded as unfettered forms of communication, but are determined by the spheres of activity in which language is used (1986, pp. 60-64, 81). Later critics have seen Bakhtin as a precursor of the newer literature on speech acts and genres (Wierzbicka, 2003, p. 457).

---

[8] Connections between speech act philosophy and diplomatic scholarship have also been examined by Pekka Henttonen (2007), but Henttonen was uninclined to challenge diplomatists' law-centred worldview, and his conclusions are rather different from those proposed here.

Other genre theorists have implicitly or explicitly made use of Bakhtin's work when proposing that 'all genres stem from speech acts' (Todorov, 1990, p. 19) or that 'acts of speech fall into genres' (Currie, 2004, p. 54; see also Bazerman, 2004, p. 309 et seq.; Post, 2013, p. 31 et seq.). Nevertheless, although some scholars have argued that genres should be treated 'in the same way as ... speech acts' (Wierzbicka, 2003, p. 149), or have used speech act theory to establish categories in a 'genre taxonomy' (Yoshioka, Herman, Yates, & Orlikowski, 2001, p. 435), it would be wrong to assume that genre categories and speech-act categories are identical. While some genres of written text, such as summonses, affidavits, or declarations of war, appear to correspond to single speech acts, most are more complex. An order for the purchase of goods can be expected to include both a (directive) request for the goods to be supplied and a (commissive) undertaking to pay for them. A typical business letter contains numerous assertive and directive speech acts, and may also include commissive, expressive, and even declarative acts. Many instances of other genres, such as office memos and email messages, comprise multiple speech acts.[9]

Although archival scholars whose work is inspired by ideas about genre have rarely discussed the theory of speech acts, it seems very relevant to their endeavours. In particular, there is an obvious affinity between understanding records through the lens of speech act theory and Fiorella Foscarini's projected reconceptualisation of records as 'social action', which derives its intellectual basis, not from Reinach or Reid, but from the North American school of rhetorical genre studies. While recognising that the world shapes the form and function of records, concepts from speech act theory can also elucidate how 'records ... organize our world' (Foscarini, 2013, unpaginated).

Contexts

Speech acts – or 'social acts', if we prefer – are always performed in contexts. As we have seen, they operate through representation and communication; communicating certain written marks or sounds counts as asserting a proposition, issuing an order, making a promise, or entering into an agreement. When this is achieved by means of written marks, the persistence of writing enables the survival of a record of the action that has been performed. But speech act philosophers have noted that communicating such marks may not count as performing an action if the circumstances are inappropriate; for example, the words of a promise or a tax demand set out in a novel do not effect a commitment or an obligation to pay. Searle (1995, pp. 54-55) proposed an explanatory formula: 'X counts as Y in context C'. The context must be suitable if the marks are to count as performing a social act. Tax demands issued by someone entitled to issue them, or the words of written promises expressed on apposite occasions, perform acts of the kind that speech act theory identifies.

Furthermore, every context has contexts of its own. The particular context in which a social act is performed – and in which a record is created – forms part of a

---

[9] Nevertheless, as Bazerman (2004, p. 320) noted, complex texts often have a single dominant action that seemingly defines their intent or purpose. A charter, for example, is likely to include a number of expressive, assertive, and directive speech acts, but these are usually seen as secondary to the declarative (or, in diplomatic terminology, dispositive) act that forms the charter's centrepiece; philosopher Daniel Vanderveken (2001, p. 255) called this a 'master' speech act.

cosmology of wider contexts, with no definable boundaries; any single aspect of context merges seamlessly into innumerable others. Acts and records operate within contexts that appear both particularised and infinite.

Archivists, of course, know that records are created in contexts. However, context remains a negotiable term, open to differing interpretations in archival literature. Some archivists equate context with the traditional understanding of the 'provenance' of records in terms of the individual, family, or organisation that created or received them (Pearce-Moses, 2005, p. 317); some seek to extend notions of both context and provenance beyond the immediate origins of records to encompass their societal framing and their histories of custody and use (Bastian, 2003, pp. 81-83; Nesmith, 2002, pp. 35-36); others see provenance as a narrow concept and affirm that context refers to a broader range of phenomena including the social, cultural, functional, and legal environments of records (Horsman, 2011, p.2; Schwartz, 2011, p. 73).

In conventional diplomatic scholarship, the pre-eminent and necessary context in which records are created is identified as a 'juridical system'. The word 'juridical' is characteristic of civil-law environments and is rarely used in Anglophone countries, but its connotations are similar to those of the English word 'legal'. Writing from a philosophical perspective, Chaim Perelman (1980, p. 168) identified the probable 'initial theses of a juridical system' as constitutional principles, laws, judiciary precedents, and 'general principles of law'. Although archivists schooled in diplomatic no longer perceive the juridical system as the sole context for record creation and have delineated an extended range of contexts (Duranti, 2010, p. 1596),[10] they have continued to affirm the pervasiveness of law (Duranti, 1998, p. 61) and the determinant influence of the juridical system on the authorship or origination of records (Duranti, 1991/2, p. 4).[11]

One reason why diplomatic scholarship assigns a pivotal role to legal or juridical systems is its abiding concern with the identification of approved forms of wording as a means of testing authenticity. Diplomatists ask whether a document employs the wording that the juridical system requires to effect a transaction or to create legally acceptable evidence of a transaction. More precisely – since the standpoint of traditional diplomatic is retrospective – they ask whether it employs a wording that

---

[10] Duranti (1997, p. 217) identified four types of context; in later work (2005, p. 27), she added a fifth. In the model proposed by Livia Iacovino (1998, p. 223), the juridical system remains predominant, and regulatory, provenancial, and documentary contexts are all subordinate to, or part of, a juridical context or system. To many archivists working outside the diplomatic tradition, these types of context are interdependent rather than hierarchical or fully distinct, and can be seen as 'forms of networks with nodes or nexuses of action, agents and relationships' (Gilliland, 2014, p. 19).

[11] In attempting to accommodate records whose juridical character may be thought doubtful, Duranti and others have sought to give the term 'juridical system' a wider meaning beyond its purely legal significance. Duranti (1991/2, p. 4) redefined it as a 'social group'; for Iacovino (2005, p. 260), the concept of a juridical system embraces professions, institutions, communities, and private social associations that issue rules or standards. However, it appears that such social groups must have law-like or binding rules, or must follow routine procedures or habits, if their records are to come within the purview of diplomatic (Duranti, 1998, pp. 44, 61). In more recent work, Duranti has sometimes (e.g., 2010, p. 1596) replaced the term 'juridical system' with 'juridical-administrative system', but the continuing emphasis on binding rules has ensured the survival of a predominantly legalistic view of records (cf. Brothman, 2011, pp. 289, 310). In similar fashion, Henttonen's (2007; 2017) studies of records and speech acts tended to assign a privileged position to records associated with formal institutional settings and rules.

the juridical system might have recognised at the time the document purports to have been created. In the sphere of diplomatic, written form is of concern only insofar as laws (or binding systems of rules that resemble laws) have taken cognizance of it, and documents whose language remains 'below the perception threshold of the law' (Vismann, 2008, p. 11) are marginal or problematic. Law, it seems, remains the context of primary interest.

In contrast, speech act philosophers make no attempt to identify a single, privileged context, and they do not claim to be able to define every aspect of context. Since context is limitless, any attempt at exhaustive classification seems predestined to fail. But speech act theory offers insights into the kinds of immediate contexts in which social acts are performed (and, if their performance involves the creation of written records, the immediate contexts in which these records come into being). It also seeks to analyse the means used to perform such acts and the contextual conditions needed for successful performance.

Social acts need not be achieved by linguistic means; they can be performed by gestures, such as raising one's arm to vote at a meeting or bid at an auction. However, language supplies conventions that allow communications among humans to be more precisely formulated and understood, and permits social acts of greater complexity. Speech act theory investigates how spoken or written language is used to perform such acts in particular contexts. When applied to writing, it can serve to explicate written records created in informal as well as formal settings, love letters as well as letters patent, and promises to mow the lawn as well as legal contracts. Instead of seeing language from the perspective of the law, it helps to situate legal and quasi-legal records within a wider consideration of language, its uses and conventions, and its embedding in society.

In examining what can be done with language, speech act philosophers are usually uninterested in regulations that seek to dictate the words used to perform specific actions. Legal systems sometimes attempt to minimise the possibility of dispute by prescribing exact wordings, but actions can also be achieved by documents whose wording is discretionary. Business letters, for example, though sometimes stigmatised by diplomatists because their written form is not juridically required, can perform acts of making promises, agreements, instructions, or permissions. Linguistic and social conventions allow these acts to be achieved; insofar as there are constraints on the wording used, they are constraints of language, not of law.

Social conditions

Some kinds of act simply require speakers or writers to know – and comply with – established linguistic practices. According to psycholinguist Charles Osgood (1979, p. 207), all known human languages are capable of expressing assertions. In Searle's words, 'to make a statement ... I need only obey the rules of language' (1979, p. 7). Many directive acts, such as making a request, are also based purely on linguistic conventions; anyone with the necessary language skills can make a request of someone else, and anyone who has learnt to put language into writing can make a request in written form. But some directive acts, such as orders and commands, require more than linguistic ability; because they operate only in situations where

one person has authority to give orders to another, they presuppose differences in social status, and their effectiveness depends not only on language but also on mutual acknowledgement of particular orderings of society. They operate under social conditions (Bourdieu, 1991, p. 73). Likewise, commissive acts such as promises require an understanding of socially-constructed notions of obligation (Hume, 1740) as well as compliance with linguistic practice.

Declaratives are perhaps the most complex acts in terms of the social conditions they presuppose. Naming ships seems to require little more than an understanding that names can be conferred on physical entities; but most declaratives operate only in societies whose members assent to intangible human institutions such as corporations, congresses, treaties, and governments, and deontic phenomena such as rights, obligations, and ownership of property. As Searle (1979, p. 18) noted, 'it is only given such institutions … that one can … bequeath one's possessions or declare war'.

The range of speech acts in use, and the extent to which employing words in particular ways counts as performing such acts, can be expected to vary from one society to another. An oft-quoted example of this diversity is the Ilongot people in the Philippines, among whom directive speech acts are commonplace, but acts of promising are unknown (Rosaldo, 1982). To a considerable degree, the contexts deemed necessary for the successful performance of a speech act may also be locally determined.

The work of genre scholars can help to elucidate many of these points. Bakhtin and members of his circle observed that each sphere of human activity and each historical period has its own repertoire of speech forms (Bakhtin, 1986, pp. 60, 64; Gardiner, 1992, p. 14), and other genre theorists have followed them in affirming the attachment of genre to time and place. Bazerman (2002, p. 342; 2004, p. 318) argued that the conditions for the genres within which speech acts are employed are likely to be far more local than Searle suggested, and that people who work together can be expected to employ distinctive sets of genres. Recognition that, within a given time period, some genres are likely to be 'widely accepted in … industrial nations' while others are specific to particular groups of people (Yates & Orlikowski, 1992, p. 304) has led genre theorists to the view that genres are enacted to realise social purposes within *communities* (Orlikowski & Yates, 1994, p. 542). Tacit understandings of generic forms are acquired through community membership (Gardiner, 1992, p. 82), and it is in communities that genre conventions are constructed, provisionally stabilised, and intermittently amended.

Although the notion of 'community' has not been immune from criticism (Foscarini, 2014, pp. 8-9; Muehlmann, 2014, pp. 582-583), it has been widely adopted. Scholarly literature abounds with references to discourse communities (Rafoth, 1988), rhetorical communities (Miller, 1994), interpretive communities (Fish, 1980), or speech communities (Gumperz, 2009). Besides genre theorists, scholars in other disciplines have adopted these terms with varying nuances of meaning and varying degrees of indebtedness to writings about genre. Ideas about communities of practice (Wenger, 1998) have been associated with genre theory

(Weldon, 2008) and occasionally with speech act theory (Wiliam, 2001). Archival scholars have written about workplace communities (Foscarini, 2014) and, with a rather different focus, communities of records or communities of memory (Bastian, 2003, pp.3-6; McKemmish, Gilliland-Swetland, & Ketelaar, 2005). Depending on the emphasis we seek, it seems that we can take our pick of the term we prefer.[12]

From a record-keeping perspective, writings about communities serve as a reminder that the contexts of records lie in localised practices as well as general rules, and that the forms invoked when actions are performed through writing are likely to depend on tacit understandings of what constitutes a social act within a given community as well as on explicit mandates or standards. However, studies of discourse communities and the like often fail to recognise that communities can endow written documents with the ability to generate and transfer rights, duties, obligations, and other deontic phenomena. Additional perspectives are needed to supply a fuller account of records and their roles in society.

Because many deontic phenomena have a legal or quasi-legal position, there should be a place here for notions of a juridical system, as diplomatists insist. The connections between records and the law are of long standing in Europe, and legal considerations must not be overlooked. However, we must be wary of attributing universal validity to concepts derived from a particular European tradition. Writing can be used to convey obligations and permissions in societies beyond the ambit of Euro-centric law; in Western societies, written promises and instructions are commonly honoured even when the law does not acknowledge or enforce them. By incorporating aspects of speech act theory, it should be possible to recognise that not all deontic phenomena are rooted in law, and that some are simply admitted by members of local communities who accept that obligations can be brought into existence through writing.[13] The theory of speech acts could thus help unite communitarian perspectives with diplomatic explanations of legal or juridical systems. A multidisciplinary approach is required if we are to develop rich intellectual frameworks for understanding records, their contexts of origin, and the ways in which they establish and underwrite social relationships.

---

[12] Some scholars emphasise the role of the community in *creating* resources; others stress its role in *using* or *interpreting* them. Some (e.g., Brown & Duguid, 1996, p. 6) also note that resources can be deployed across community boundaries. There are surprisingly few scholarly references to 'speech act communities', but a few can be found (Eppinger, 2013; Fish, 1980, p. 244). In other fields, particularly those closer to the 'hard' sciences, broadly similar concepts have sometimes been expressed using the word 'domain' rather than 'community': biologist Humberto Maturana's notion of a consensual domain has been applied to linguistic activity (Winograd & Flores, 1986, p. 49), and writers influenced by ideas about information systems or computer-science ontologies often refer to domains of discourse (Beynon-Davies, 2009, pp. 80-81; cf. Lemieux, 2014, pp. 76-77).

[13] Notions of (what diplomatists call) *competence* – 'the authority and capacity of accomplishing an act' (Duranti, 1998, p. 89) – are also found in speech act theory, where they are variously known as *entitlement* (Austin, 1962, pp. 34-35), *capacity* (Austin, 1962, p. 23), or *status* (Searle, 1979, pp. 5-7). A captain can issue instructions to a sergeant, but not to a colonel; I can bequeath my possessions, but cannot bequeath yours. Duranti (1998, pp. 90-91) noted that, where written acts are concerned, a number of different competences may be involved; for example, competence to act and competence to issue or establish the forms of a document may rest with different persons. But whereas diplomatic restricts the competence to act to physical and juridical persons (i.e., 'persons' formally recognised by the law), speech act theory imposes no such restriction. Informal groups of people, ad-hoc teams, and unincorporated associations without legal status can ask and answer questions; in many communities, they can also make promises and agreements. If they perform these acts in writing, they create records.

Postmodernism and 'the performative'

In the late 20th century, speech act theory engaged the attention of many philosophers in what may loosely be called postmodernist schools of thought. Jean-François Lyotard (1984) adopted Austin's concept of performative language in developing his ideas about modes of discourse, and used notions of performativity to explain the legitimisation of knowledge in an era he saw as characterised by disbelief in metanarratives. For Paul Ricoeur (1971, pp. 537-544), an 'action-event' resembles a speech act, but both written texts and 'action-events' can be dissociated from the conditions of their production and from the intentions of authors or actors. Michel Foucault, too, although often seeking to distance himself from Austin and Searle, frequently drew on their work in his analysis of discursive practices and the power of language (Foucault, 1972; Mottier, 2008; cf. Dreyfus & Rabinow, 1983, p. 46).

Among these philosophers, the most sustained engagement with speech act theory was that of Jacques Derrida. Initially drawn to Austin's thinking because it presented 'the performative' as non-referential, Derrida (1988, pp. 13-14) observed that Austin had seemingly shattered the notion that oral or written communication was confined to 'the transference of … semantic content … dominated by an orientation toward truth'. However, Derrida (1988, pp. 14, 137) repudiated Austin's assumptions about the conscious presence of a speaker (or writer) and a definable context in which a speech act is performed, and argued that the finiteness of contexts is 'never secured'. He insisted on what he called the 'iterability' of utterances: their potential for infinite repetition and citation, which – according to Derrida (1988, pp. 17-18) – calls into question the possibility of a 'pure' performative speech act.

Much of Derrida's later writing was formulated from within his critique of Austin's work (Ortiz-Robles, 2005). This was evident, for example, when he returned to the ideas surrounding his well-known affirmation that 'the archivization produces as much as it records the event' (Derrida, 1996, p. 17), in his comments on the terrorist attacks of September 2001. For Derrida, an event of this kind 'is made up of the "thing" itself … and the impression … that is given' when it is inscribed or recorded; 'inscription produces a new event, thereby affecting the presumed primary event it is supposed to retain … [and] archive' (Borradori, 2003, p. 89; Derrida, 2002, p. 113). In themselves, these thoughts were not greatly different from his earlier ideas, but when he came to describe the processes of inscription or 'saying the event', he incorporated concepts of 'the performative' derived from his interpretation of Austin; he observed that there is 'a dimension of saying the event that overtly presents itself as performative' (Derrida, 2007, p. 447). Moreover, although 'the performative says and produces the event that it speaks of', it also – according to Derrida (2007, p. 460) – attempts to 'neutralize' and 'reappropriate' the event. In embracing notions of performativity, Derrida reinvented them to suit his own ends.

Postmodernist thinkers are often ambivalent about speech act theory. While welcoming its rejection of the belief that the sole purpose of language is to communicate facts about the world, they are usually suspicious of its systematising tendencies and its claim that utterances can perform definable actions or change

the world in a determinate manner.[14] When applied to records, speech act theory reminds us that records are linked to acts performed on specific occasions; it perplexes those critics who insist that there can be no distinct originary moments. Whereas the Foucauldian tradition maintains that discourse or text must be understood without reference to the intentions of speakers or writers, speech act theory invites us to recognise an original 'first writing' that effects an action. Words and phrases are undoubtedly iterable, but 'performances of ... speech acts (whether written or spoken) are ... datable singular events in particular historical contexts' (Searle, 1977, p. 208).

Viewed in this light, the singularity of an event implicates, and is implicated by, the singularity of its context. In Bakhtin's words, while 'each utterance is filled with echoes ... of other utterances to which it is related' (1986, p. 91), every utterance also has an 'unrepeatable individual context' (1986, p. 88). From Derrida's perspective, however, contexts can never be self-identifying, and attempts to fix the contexts of utterances are always political (1988, p. 136). Other commentators have responded to Derrida's concerns by emphasising that the contexts in which speech acts are performed are not autonomous but are, to some degree, a matter of mutual agreement and shared assumptions. A written utterance can establish a permission or obligation because the members of a community jointly assume that such phenomena can be generated through writing in agreed contexts. Where shared assumptions of this kind are absent, a speech act is unlikely to be successful (Fish, 1999, p.70). Moreover, speech acts themselves create and sustain the social settings in which speech acts occur; contexts and activities are mutually constitutive (Dourish, 2004, p. 28; Fish, 1980, p. 216). This interdependency enables records to function and underpins our ability to use them to perform actions in the world.

Information and representation

Finally, let us return to the topic of 'information', with which this essay began. Our reflection on speech act theory leads us away from the ingenuous perception of records as mere purveyors of facts, and underscores their intimate association with contextualising and contextualised action; in the light of this, it seems clear that we must relinquish any beliefs that records are simply 'information objects' or that records and information are near-identical. Where, then, might we accommodate ideas about information in a view of records that takes account of speech act philosophy?

Most obviously, information can be sought in records that Searle's taxonomy labels *assertive*. Many records of this kind merely set out a proposition, but occasionally we find records whose creators explicitly claim to be 'informing' others. In the following example dating from ca.1065 (given here in modern English translation,

---

[14] See, for example, Derrida (2002) and (2007). Although in these writings much of Derrida's interest was focused on what speech act theorists call *assertive* acts, such as journalistic reporting of terrorist attacks, he also gave voice to ideas about other performative speech acts. He readily agreed that 'when I make a promise, I'm not saying an event; I'm producing it by my commitment; ... "I promise" is a saying ... that produces the event' (2007, pp. 446, 458). But he was more hesitant about 'harder' forms of commitment. Ideally, perhaps, he would have liked to dismiss contracts, 'civil status', and the law – emblems of a world in which speech acts generate deontic phenomena – as 'so many fables' (2002, p. xvi). Insofar as performatives succeed in creating 'what is ... held to be ... juridically incontestable public truth', this – in Derrida's view – should be problematised or condemned as 'performative violence' (2002, pp. 51, 231).

from Harmer, 1952, p. 282), the record creator is an English king, who writes formally to his subordinates:

> *King Edward sends greetings ... I inform you that I have given to Bishop Giso the land at Wedmore ...*

An even earlier (and less formal) example is a business letter written by a cattle overseer in ancient Egypt, about 3000 years ago (translation from Wente & Meltzer, 1990, p. 31):

> *This is a missive to inform my lord that I am carrying out ... every assignment that has been charged to me. ... I am writing to inform my lord that a message should be sent ...*

Modern equivalents using the word 'inform' can easily be imagined:

> *Dear Ms Bloggs,*
> *I write to inform you that the Board has decided ...*

Even when creators of 'assertive' records do not use a word such as 'inform', the propositions they state ('We have sold the property'; 'I have completed the task'; 'The Board has resolved to issue new shares') could perhaps be seen as information that they wish to convey to readers of the records concerned. Advocates of the view that records 'contain information' rarely discuss propositions, but what they mean by 'information' would seem to be the propositions that record creators state.

Suppose, for example, that a Board Secretary writes:

> *Dear Ms Bloggs,*
> *The Board has agreed the terms of the new share issue ...*

Here, a proposition is the only explicit content of the record. The Secretary is stating a proposition about what the Board has agreed to do, but the record appears to be conveying autonomous information, because the act of making a statement remains implicit and is concealed from the reader.

However, any concept of information must take account of the possibility of misinformation. It is legitimate to ask how far records can be said to contain or provide information, if some of the propositions set out in them seem inaccurate, mistaken, biased, or distorted to tell of an ideal past. It could be argued that the cattle overseer supplied his manager with the *information* that he was carrying out every task assigned to him, even if several of his tasks had actually remained untouched. But it is also possible to contend that, when propositions are fallacious, they cannot appropriately be labelled as information. For Luciano Floridi (2004, pp. 42-46), a 'general definition of information' does not require information to be truthful; truthfulness is a condition only of a 'special definition of information'. This distinction remains characteristic of Floridi's work. Other philosophers have considered the issue without resolving it. Some speculate that information could be either true or false; others insist that falsehoods cannot be information (Hennig, 2014, pp.251-252).

Truth, of course, is itself a contested notion; not every commentator would accept that propositions can be characterised as definitively true or false. A less foundational stance suggests that propositions can only express perceptions of the world, and that stating a proposition entails consciously or unconsciously selecting one way of representing the matter to which the proposition refers, while excluding others that could be equally plausible. Stating a proposition is a social practice that necessarily reduces complex realities to manageable verbal forms.

Whenever record creators assert propositions, a cautious approach may lead us to conclude that those propositions are open to dispute or at least to variable interpretation. The studies cited earlier in this chapter, which seek to demonstrate that medical records or social work files are tendentious, all focus on records that depend on the assertion of propositions about past events. Because these records are almost always constructed at an interval of time after − and often also at a place distant from − the actions and events they describe, they are liable to distortion or bias in favour of the interests of their creators or the organisations for which their creators work.

Concerns about the reliability of propositions asserted by records creators need not be limited to records that report retrospectively on past events. Making a statement of any kind entails producing a particular representation of the way things were, are, or are thought likely to be, and it cannot be a neutral practice. Of course, a simple assertion of, for example, a cost estimate offers less scope for improvisation than more discursive or creative forms of record, but all assertions involve record creators in choosing to present their message in a particular way. Even when institutions or legal systems attempt to impose regulated vocabularies, there is almost always space left for authorial choice, which in turn may lead us to question the objectivity of the propositions asserted in a record.

Almost certainly, however, when we encounter such propositions, we will be less inclined to deny that the creator of the record has asserted them. Although we may choose not to believe the Egyptian overseer's claim that he was working on each of his assignments, we will be unlikely to doubt that he asserted a proposition to that effect. Likewise, when we read a file of job application forms, we may question the veracity of statements made by individual applicants, but such questioning does not diminish our understanding of the file as a record of the statements that were made during the application process.[15]

Seeing records of this kind as 'information' is less fruitful than seeing them as representations of propositions asserted and of the acts of asserting them. In contrast to the popular view that information wants to be 'free' and enjoy 'a life of

---

[15] We can draw very similar conclusions about the metadata, or descriptions, that records' custodians or other agents create. Records professionals seek to secure the contextualisation of records by surrounding them with appropriate metadata, but metadata are not exempt from error, distortion, social constraints, or human judgement. Metadata are created at specific moments of time; they always have contexts of their own, and they can be understood as records of propositions that were asserted in the course of a descriptive process. Although metadata are commonly characterised as 'data about data', it is equally possible − and, from a speech-act viewpoint, more productive − to see them as records of assertions made about other records, entities, or relationships.

its own' (Losey, 2015), a perspective drawn from speech act philosophy reminds us that information is always linked to its contexts of origin and to the actions of individuals in society. As historian John Tosh (2015, p. 108) noted, records are not mere 'testimonies of events "out there", but … parts of a process' of acting and recording that can itself be the subject of inquiry. The propositions that people state cannot be wholly separated from the people who state them or the systems of representation in which they are entangled.

Revisiting 'information' from a speech act perspective

The class of *assertive* records includes (but is not limited to) records in which the creator formally asserts a proposition describing a juridical action completed in the past. Diplomatists, who call such records 'probative', insist that they are purely evidentiary and that only a dispositive record can function performatively. But assertions do not merely convey evidence or information; they have their own performative characteristics. The formal document in which King Edward asserted that he had given land to the bishop cannot simply be equated with the information that the king made this gift, or even with the information that he asserted that he had made it; rather, the issuing of the document performed the king's act of assertion. Diplomatists analysing this document are likely to assume that the only action of significance is the royal gift of the land; from the standpoint of diplomatic, this assumption seems justified, because diplomatists understand that 'probative' records were and are created to provide evidence of actions of this kind, and because legal systems recognise the records' evidential role. From a speech-act perspective, however, a record of this kind performs an action of its own. As Derrida (2002, p. 113) observed, the inscription produces a new event.

Diplomatic analysis reflects diplomatists' legalistic view of records and responds to the needs of those users (including lawyers, litigants, and legal historians) who view records in a similar fashion. It appears less suited to many humanist contexts of study, where users may wish to look at assertive records in alternative ways and may perhaps be less interested in propositions than in the activity of asserting them or the circumstances in which they were asserted. The law looks for verification, but an act of asserting cannot be assessed in terms of truth or falsehood, qualities that could apply only to the propositions stated or asserted. Assertive records display a complex performativity, which cannot be encompassed by perceiving them simply as probative or informational.

Of course, *directive* records, such as a summons to appear in court or an invitation to give a lecture, are not statements of propositions; they are instructions or requests. Similarly, a *commissive* record, such as a promissory note, or a *declarative* record, such as a will, does not merely state that the creator of the record has made a promise or bequeathed a watch; when the note is communicated or the will is proved, the record creator performs the act of promising or bequeathing. We may perhaps doubt whether an act of promising is sincere, but when we encounter a record in which the creator says 'I promise …' we can hardly doubt that a promise has been made (Yeo, 2010, pp. 105-106). Like acts of asserting, an act of promising, ordering, or declaring cannot be judged using criteria of truth or falsehood, or of the accuracy or inaccuracy of information. When such an act is performed, word and

deed belong together; the record is intimately involved in the action that it represents. Although archivists have sometimes claimed that records are 'information about ... action' (Upward, Reed, Oliver, & Evans, 2013, p. 48; my italics), speech act philosophy impels us to the view that every record is an *instrument of* action. Records are not mere descriptions of actions or events; they are part of the way in which business is conducted and lives are lived.

Rather than seeing information as content embedded in records at the point of their creation, we may find it more profitable to associate information with *uses* of records in their later life. Most obviously, users can employ a record to acquire information about the action that its creator performed or the propositions that its creator stated. They can use it in this way because the record is persistent; it continues to represent the action after the action has been performed. If they have a 'directive' record that represents instructions I have given, or a 'commissive' record that represents a contract I have agreed, they can use it to gather information about my instructions or my agreement. If they have an 'assertive' record that represents my action of stating a proposition, they can use it to gather information both about the proposition and about my action in stating it.

However, they can also use records to acquire information about other topics: the social contexts in which a record was made or kept, the record-making and record-keeping practices that were employed, the modes of life and thought of the people who created it or are mentioned within it, the resources that were available to these people, and much else besides. A user can garner a vast range of information from a record, even when informing readers of a proposition was not its original purpose. When users examine a record in which a writer stated a proposition about the world, the information they can derive need not be confined to the proposition that the writer sought to convey.

There is little to be gained from suggesting that records comprise information but can also be used to garner *other* information. Instead, an understanding of records founded on speech act philosophy allows us to see information as an affordance that arises from engagement with records. Information, in this sense, is bound to circumstances; it depends on the intellectual processes applied by the user and the mental frameworks that users bring to the interpretation of the record, as well as on the record's content and structure.[16] A user can acquire different information from a single record in different episodes of use. The information that one user derives from this record may be very different from the information that another user derives from it. Each user 'may see ... new information that no one else has seen before' (Latham, 2011, p. 13). It may also be possible for a user to acquire similar information from each of several different records.

Moreover, information is just one of many affordances obtainable from records. As Terry Eastwood (1993, p. 112) noted, records 'frequently suffer from being transformed into mere sources of information, when they are in fact much more than that'. Records can afford, not only information, but evidence, accountability, and senses of personal or social identity, as well as emotions, ideas, inspirations, or

---

[16] According to Jonathan Furner (2014, p. 166), this was the view of philosopher Agnès Lagache: 'Information is in the receiver. ... Information arises from his reading. He creates it.'

guidance for future action. Their persistence allows them to sustain or corroborate individual or communal memories. They also often have aesthetic qualities and symbolic connections with particular people, institutions, or places. Their richness of affordances transcends any single aspect of their use (Yeo, 2007, pp. 330-331).

Conclusion

In recent years, 'information' has become something of a buzzword. Its importance is constantly affirmed both by governments and by the popular media, which often promote it as a key to transparency, democratic freedoms, and economic success. Records professionals have been attracted by this message and have frequently sought to adopt an information agenda. To many archivists and records managers, emphasis on information seems to offer a professional image in tune with current developments in the wider world, and the perceived association between information and the emerging digital realm gives it a further aura of desirability and prestige.

However, much recent writing about information has been driven by commercial or technological approaches that take little account of the centrality of human agency and social context. We cannot assume that records offer 'the unvarnished facts, the raw data, the actual measurements, the ... real information' (Hamm, 2011, p. 44) demanded by authors of current business textbooks, or the single 'source of truth' (Kosur, 2015; Roon, 2016) sought by many data analysts and promoters of blockchain technologies. Even when records appear to be purely factual, their content depends on utterances that are circumstantially produced. Users who are aware that 'there is ... a good deal of game playing ... in the record production business' (Van Maanen & Pentland, 1994, p. 58) will use this knowledge to assist them in formulating and assessing information from the records they encounter.

Looking at records through the lens of speech act theory can help us gain a richer understanding of 'the record production business'. While a speech-act view of records does not deny that a record may undergo many adventures in its later life, or that these adventures can be perceived as incessant processes of recontextualisation (Caswell, 2016; McKemmish, 2001), its primary concern is with the moment of inscription and the context in which a record first comes into being. At that contextual moment, what takes place is a matter of action, not a matter of information. Whatever may be our motive for *keeping* records or our comprehension of *using* them, their *creation* is necessarily performative.

Speech act theory reminds us that records are not mere information objects or containers of facts, but it also affirms that records do not simply dissolve into interpretation. At the point of inscription, a record and an assertive, directive, commissive, or declarative action are interlinked. In this sense, records have a specific social identity; they are integral parts of the actions they represent.

Literature

Austin, J.L. (1962). *How to do things with words*. Oxford: Clarendon Press.

Bailey, S. (2008). *Managing the crowd: Rethinking records management for the Web 2.0 world*. London: Facet Publishing.

Bakhtin, M.M. (1986). *Speech genres and other late essays*. Austin: University of Texas Press.

Bastian, J.A. (2003). *Owning memory: How a Caribbean community lost its archives and found its history*. Westport: Libraries Unlimited.

Bátori, Z. (2015). Photographic deception. *Proceedings of the European Society for Aesthetics, 7*, 68-78.

Bazerman, C. (2002). *The languages of Edison's light.* Cambridge, MA: MIT Press.

Bazerman, C. (2004). Speech acts, genres, and activity systems. In C. Bazerman & P. Prior (Eds.), *What writing does and how it does it*. Mahwah: Lawrence Erlbaum.

Bazerman, C. (2012). The orders of documents, the orders of activity, and the orders of information. *Archival Science, 12*, 377-388.

Berg, M. (1996). Practices of reading and writing: The constitutive role of the patient record in medical work. *Sociology of Health and Illness, 18*, 499-524.

Beynon-Davies, P. (2009). *Business information systems*. Basingstoke: Palgrave Macmillan.

Boles, F., & Greene, M. (2001). Confusing the bun for the burger: Rehabilitating the role of content in the archival context. *Schweizerische Zeitschrift für Geschichte, 51*, 424-447.

Borradori, G. (2003). *Philosophy in a time of terror: Dialogues with Jürgen Habermas and Jacques Derrida*. Chicago: University of Chicago Press.

Bourdieu, P. (1991). *Language and symbolic power*. Cambridge (UK): Polity Press.

Bowman, S. (2013). Impact of electronic health record systems on information integrity: Quality and safety implications. *Perspectives in Health Information Management, 10*. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3797550/

Boyle, L. (1992). Diplomatics. In J.M. Powell (Ed.), *Medieval studies: An introduction* (2nd ed.). Syracuse: Syracuse University Press.

Brothman, B. (2002). Afterglow: Conceptions of record and evidence in archival discourse. *Archival Science, 2*, 311-342.

Brothman, B. (2011). Designs for records and recordkeeping: Visual presentation in diplomatics, the record continuum, and documentation strategy. In T. Cook (Ed.), *Controlling the past: Documenting society and institutions*. Chicago: Society of American Archivists.

Brown, B.A.T. (2001). Unpacking a timesheet: Formalisation and representation. *Computer Supported Cooperative Work, 10*, 293-315.

Brown, J.S., & Duguid, P. (1996). The social life of documents. *First Monday, 1*(1). Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/466/820

Caswell, M. (2016). 'The archive' is not an archives: acknowledging the intellectual contributions of archival studies. *Reconstruction, 16*(1). Retrieved from http://reconstruction.eserver.org/Issues/161/Caswell.shtml

Clarke, S. (2009). Crowded out: Records management and the Web 2.0 phenomenon. *Archives and Manuscripts, 37*, 118-133.

Cochran, N., Gordon, A.C., & Krause, M.S. (1980). Proactive records: Reflections on the village watchman. *Knowledge: Creation, Diffusion, Utilization, 2*, 5-18.

Cook, T. (1994). Electronic records, paper minds: The revolution in information management and archives in the post-custodial and post-modernist era. *Archives and Manuscripts, 22*, 300-328.

Cook, T. (2001). Fashionable nonsense or professional rebirth: Postmodernism and the practice of archives. *Archivaria, 51,* 14-35.

Cooren, F. (2004). Textual agency: How texts do things in organizational settings. *Organization, 11*, 373-393.

Cornelius, I. (2014). Epistemological challenges for information science. In F. Ibekwe-SanJuan & T.M. Dousa (Eds.), *Theories of information, communication and knowledge*. Dordrecht: Springer.

Currie, G. (2004). *Arts and minds*. Oxford: Clarendon Press.

Derrida, J. (1988). *Limited inc*. Evanston: Northwestern University Press.

Derrida, J. (1996). *Archive fever: A Freudian impression*. Chicago: University of Chicago Press.

Derrida, J. (2002). *Without alibi*. Stanford: Stanford University Press.

Derrida, J. (2007). A certain impossible possibility of saying the event. *Critical Inquiry, 33*, 441-461.

Doty, K.L., & Hiltunen, R. (2009). Formulaic discourse and speech acts in the witchcraft trial records of Salem, 1692. *Journal of Pragmatics, 41*, 458-469.

Dourish, P. (2004). What we talk about when we talk about context. *Personal and Ubiquitous Computing, 8*, 19-30.

Dreyfus, H.L., & Rabinow, P. (1983). *Michel Foucault: Beyond structuralism and hermeneutics*. Chicago: University of Chicago Press.

Duranti, L. (1991/2). Brothman on authorship: The diplomatic perspective. *Archivaria, 33*, 4-5.

Duranti, L. (1997). The archival bond. *Archives and Museum Informatics, 11*, 213-218.

Duranti, L. (1998). *Diplomatics: New uses for an old science*. Lanham: Scarecrow Press.

Duranti, L. (2005). *The long-term preservation of authentic electronic records: Findings of the InterPARES project*. San Miniato: Archilab.

Duranti, L. (2010). Diplomatics. In M.J. Bates & M.N. Maack (Eds.), *Encyclopedia of library and information sciences* (3rd ed.). Boca Raton: CRC Press.

Duranti, L., & Preston, R. (Eds.) (2008). *International research on permanent authentic records in electronic systems (InterPARES 2)*. Padua: Università di Padova.

Eastwood, T. (1993). How goes it with appraisal? *Archivaria, 36*, 111-121.

Eastwood, T. (2017). A contested realm: The nature of archives and the orientation of archival science. In H. MacNeil & T. Eastwood (Eds.) , *Currents of archival thinking* (2nd ed.). Santa Barbara: Libraries Unlimited.

Eppinger, M. (2013). Sages, savages, and other speech act communities. *Saint Louis University Law Journal, 57*, 407-428.

Ferraris, M. (2013). *Documentality: Why it is necessary to leave traces*. New York: Fordham University.

Financial Conduct Authority (ca.2015). *Records management principles*. Retrieved from https://www.fca.org.uk/static/documents/foi/records-management-principles.pdf

Fish, S. (1980). *Is there a text in this class?* Cambridge, MA: Harvard University Press.

Fish, S. (1999). Reflections on Austin and Derrida. In H. Aram Veeser (Ed.), *The Stanley Fish reader*. Oxford: Blackwell.

Floersch, J. (2000). Reading the case record: The oral and written narratives of social workers. *Social Service Review, 74*, 169-192.

Floridi, L. (2004). Information. In L. Floridi (Ed.), *The Blackwell guide to the philosophy of computing and information*. Oxford: Blackwell.

Foscarini, F. (2013). Record as social action: Understanding organizational records through the lens of genre theory. *Information Research, 18*(3). Retrieved from http://InformationR.net/ir/18-3/colis/paperC08.html

Foscarini, F. (2014). A genre-based investigation of workplace communities. *Archivaria, 78*, 1-24.

Fotion, N. (2000). *John Searle*. Teddington: Acumen.

Foucault, M. (1972). *The archaeology of knowledge*. London: Tavistock Publications.

Furner, J. (2014). Information without information studies. In F. Ibekwe-SanJuan & T.M. Dousa (Eds.), *Theories of information, communication and knowledge*. Dordrecht: Springer.

Gardiner, M. (1992). *The dialogics of critique: M. M. Bakhtin and the theory of ideology*. London: Routledge.

Gilliland, A.J. (2014). Reconceptualizing records, the archive and archival roles and requirements in a networked society. *Knygotyra, 63*, 17-34.

Gitelman, L. (Ed.) (2013). *'Raw data' is an oxymoron*. Cambridge, MA: MIT Press.

Gumperz, J.J. (2009). The speech community. In A. Duranti (Ed.), *Linguistic anthropology*. Chichester: Wiley-Blackwell.

Hamm, J. (2011). *Unusually excellent: The necessary nine skills required for the practice of great leadership*. San Francisco: Jossey-Bass.

Hancher, M. (1979). The classification of cooperative illocutionary acts. *Language in Society, 8*, 1-14.

Hanks, P. (2015). *Propositional content*. Oxford: Oxford University Press.

Harmer, F.E. (1952). *Anglo-Saxon writs*. Manchester: Manchester University Press.

Hennig, B. (2014). Documents: Fillers of informational gaps. *The Monist, 97*, 246-255.

Henttonen, P. (2007). *Records, rules and speech acts*. Tampere: Tampere University Press.

Henttonen, P. (2017). Looking at archival concepts and practice in the light of speech act theory. In A.J. Gilliland, S. McKemmish, & A.J. Lau (Eds.), *Research in the Archival Multiverse*. Clayton, VIC: Monash University Publishing. http://dx.doi.org/10.26530/OAPEN_628143

Horsman, P. (2011). *Wrapping records in narratives: Representing context through archival description*. Retrieved from https://www.its-arolsen.org/fileadmin/user_upload/Metanavigation/Aktuelles/2011_PDF/04_Horsman_text.pdf

Hume, D. (1740). *A treatise of human nature, vol.3*. London: Thomas Longman.

Iacovino, L. (1998). The nature of the nexus between recordkeeping and the law. *Archives and Manuscripts, 26*, 216-246.

Iacovino, L. (2005). Recordkeeping and juridical governance. In S. McKemmish, M. Piggott, B. Reed, & F. Upward (Eds.), *Archives: Recordkeeping in society*. Wagga Wagga: Charles Sturt University.

ISO15489-1:2016. *Records management. Part 1: Concepts and principles*.

Jordan, R.L., Warren, W.D., & Walt, S.D. (2000). *Negotiable instruments, payments, and credits* (5th ed.). St Paul: Foundation Press.

Kosur, J. (2015). *CFOs need to define a 'single source of truth' for all of their business needs*. Retrieved from http://uk.businessinsider.com/cfos-need-to-define-the-single-source-of-truth-2015-10

Kurzon, D. (1986). *It is hereby performed: Explorations in legal speech acts*. Amsterdam: John Benjamins.

Latham, K.F. (2011). Medium rare: Exploring archives and their conversion from original to digital. Part two: The holistic knowledge arsenal of paper-based archives. *LIBRES 21*(1). Retrieved from http://libres-ejournal.info/wp-content/uploads/2014/06/Vol21_I1_Latham__EssOp_LIBRES21n1.pdf

Lee, R.M. (1988). Bureaucracies as deontic systems. *ACM Transactions on Office Information Systems, 6*, 87-108.

Lemieux, V.L. (2014). Toward a 'third order' archival interface: Research notes on some theoretical and practical implications of visual explorations in the Canadian context of financial electronic records. *Archivaria, 78*, 53-93.

Lloyd, S.S., & Rissing, J.P. (1985). Physician and coding errors in patient records. *Journal of the American Medical Association, 254*, 1330-1336.

Losey, R. (2015). *Information governance v search: The battle lines are redrawn*. Retrieved from http://e-discoveryteam.com/2015/02/08/information-governance-v-search-the-battle-lines-are-redrawn/

Lynch, M. (1993). *Scientific practice and ordinary action*. Cambridge (UK): Cambridge University Press.

Lyotard, J.-F. (1984). *The postmodern condition: A report on knowledge*. Manchester: Manchester University Press.

McCleary, R. (1977). How parole officers use records. *Social Problems, 24*, 576-589.

McGonigle, D., & Mastrian, K. (Eds.) (2012). *Nursing informatics and the foundation of knowledge*. Burlington: Jones & Bartlett.

McKemmish, S. (2001). Placing records continuum theory and practice. *Archival Science, 1*, 333-359.

McKemmish, S., Gilliland-Swetland, A., & Ketelaar, E. (2005). 'Communities of memory': pluralising archival research and education agendas. *Archives & Manuscripts, 33*, 146-174.

Mead, G.H. (1934). *Mind, self, and society*. Chicago: University of Chicago Press.

Miller, C.R. (1994). Rhetorical community: The cultural basis of genre. In A. Freedman & P. Medway (Eds.), *Genre and the new rhetoric*. London: Taylor & Francis.

Mottier, V. (2008). Metaphors, mini-narratives and Foucauldian discourse theory. In T. Carver & J. Pikalo (Eds.), *Political language and metaphor*. Abingdon: Routledge.

Muehlmann, S. (2014). The speech community and beyond. In N.J. Enfield, P. Kockelman, & J. Sidnell (Eds.), *The Cambridge handbook of linguistic anthropology*. Cambridge (UK): Cambridge University Press.

Nesmith, T. (2002). Seeing archives: Postmodernism and the changing intellectual place of archives. *American Archivist, 65*, 24-41.

Oliver, G. (2017). Managing records in current recordkeeping environments. In H. MacNeil & T. Eastwood (Eds.), *Currents of archival thinking* (2nd ed.). Santa Barbara: Libraries Unlimited.

Orlikowski, W., & Yates, J. (1994). Genre repertoire: The structuring of communicative practices in organizations. *Administrative Science Quarterly, 39*, 541-574.

Ortiz-Robles, M. (2005). Being Jacques Derrida. *Postmodern Culture, 15*(3). Retrieved from http://pmc.iath.virginia.edu/text-only/issue.505/15.3ortiz-roblez.txt

Osgood, C.E. (1979). What is a language? In D. Aaronson & R.W. Rieber (Eds.), *Psycholinguistic research*. Hillsdale: Lawrence Erlbaum.

Pearce-Moses, R. (2005). *A glossary of archival and records terminology.* Chicago: Society of American Archivists.

Pember, M., & Cowan, R.A. (2009). *Where is the record we have lost in information?* Retrieved from https://core.ac.uk/download/files/343/11233465.pdf

Perelman, C. (1980). *Justice, law, and argument*. Dordrecht: Reidel.

Poirier, S. (1999). Voice structure, politics, and values in the medical narrative. *HEC Forum, 11*, 27-37.

Post, M. (2013). *Speech acts and speech genres: An axiological linguistics perspective*. Wroclaw: Wyższa Szkoła Filologiczna.

Rafoth, B.A. (1988). Discourse community: Where writers, readers, and texts come together. In B.A. Rafoth & D.L. Rubin (Eds.), *The social construction of written communication*. Norwood: Ablex.

Reid, T. (1788). *Essays on the active powers of the human mind*. Edinburgh: John Bell.

Reinach, A. (1913). Die apriorischen grundlagen des bürgerlichen rechtes. *Jahrbuch für Philosophie und Phänomenologische Forschung, 1*, 685-847.

Ricoeur, P. (1971). The model of the text: Meaningful action considered as a text. *Social Research, 38*, 529-562.

Roon, M. (2016). *Blockchains: A new source of truth*. Retrieved from https://www.linkedin.com/pulse/blockchains-new-source-truth-micha-roon

Rosaldo, M.Z. (1982). The things we do with words: *Ilongot* speech acts and speech act theory in philosophy. *Language in Society, 11*, 203-237.

Saffady, W. (2011). *Records and information management: Fundamentals of professional practice* (2nd ed.). Overland Park: ARMA International.

Schwartz, J.M. (2011). The archival garden: Photographic plantings, interpretive choices, and alternative narratives. In T. Cook (Ed.), *Controlling the past: Documenting society and institutions*. Chicago: Society of American Archivists.

Searle, J.R. (1969). *Speech acts: An essay in the philosophy of language.* Cambridge (UK): Cambridge University Press.

Searle, J.R. (1977). Reiterating the differences: A reply to Derrida. *Glyph, 1*, 198-208.

Searle, J.R. (1979). *Expression and meaning: Studies in the theory of speech acts*. Cambridge (UK): Cambridge University Press.

Searle, J.R. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge (UK): Cambridge University Press.

Searle, J.R. (1995). *The construction of social reality.* New York: Free Press.

Searle, J.R. (1999). *Mind, language and society*. New York: Basic Books.

Searle, J.R. (2008). *Philosophy in a new century*. Cambridge (UK): Cambridge University Press.

Smith, B. (2008). Searle and De Soto: The new ontology of the social world. In B. Smith, D.M. Mark, & I. Ehrlich (Eds.), *The mystery of capital and the construction of social reality*. Chicago: Open Court.

Smith, B. (2014). Document acts. In A. Konzelmann-Ziv & H.B. Schmid (Eds.), *Institutions, emotions, and group agents*. Dordrecht: Springer.

Stair, R., Moisiadis, F., Genrich, R., & Reynolds, G. (2011). *Principles of information systems*. Melbourne: Cengage.

State Records New South Wales (2004). *Standard on full and accurate records*. Retrieved from http://arp.nsw.gov.au/sites/default/files/Standard%20No%20%207%20-%20Full%20and%20Accurate%20Records.pdf

Suchman, L. (1994). Do categories have politics? *Computer Supported Cooperative Work, 2*, 177-190.

Sundqvist, A. (2011). Documentation practices and recordkeeping: a matter of trust or distrust? *Archival Science, 11*, 277-291.

Thomas, D., Fowler, S., & Johnson, V. (2017). *The silence of the archive*. London: Facet Publishing.

Todorov, T. (1990). *Genres in discourse*. Cambridge (UK): Cambridge University Press.

Tosh, J. (2015). *The pursuit of history* (6th ed.). Abingdon: Routledge.

Trace, C.B. (2016). Ethnomethodology: Foundational insights on the nature and meaning of documents in everyday life. *Journal of Documentation, 72*, 47-64.

Trombley, S. (2016). Self-reinvention: The next step for future-forward information professionals. *IRMS Bulletin, 194*, 50-51.

Underwood, W. (2008). *Recognizing speech acts in presidential e-records*. Retrieved from http://perpos.gtri.gatech.edu/publications/TR%2008-03.pdf

Upward, F., Reed, B., Oliver, G., & Evans, J. (2013). Recordkeeping informatics: Re figuring a discipline in crisis with a single minded approach. *Records Management Journal, 23*, 37-50.

Van Maanen, J., & Pentland, B. (1994). Cops and auditors: The rhetoric of records. In S.B. Sitkin & R.J. Bies (Eds.), *The legalistic organization*. Thousand Oaks: Sage.

Vanderveken, D. (2001). Illocutionary logic and discourse typology. *Revue Internationale de Philosophie, 55,* 243-259.

Vismann, C. (2008). *Files: Law and media technology*. Stanford: Stanford University Press.

Von Wright, G.H. (1951). Deontic logic. *Mind, 60,* 1-15.

Weldon, P.R. (2008). *Genre and change: Genre theory and language change in a business community of practice*. Saarbrücken: Müller.

Wenger, E. (1998). *Communities of practice: Learning, meaning and identity*. Cambridge (UK): Cambridge University Press.

Wente, E.F., & Meltzer, E.S. (Eds.) (1990). *Letters from ancient Egypt*. Atlanta: Scholars Press.

Whittaker, S., Laban, R., & Tucker, S. (2006). Analysing meeting records: An ethnographic study and technological implications. *Lecture Notes in Computer Science, 3869,* 101-113.

Wierzbicka, A. (2003). *Cross-cultural pragmatics: The semantics of human interaction (2nd ed.)*. Berlin: Mouton de Gruyter.

Wiliam, D. (2001). *How to do things with assessments: Illocutionary speech acts and communities of practice*. Retrieved from http://www.nottingham.ac.uk/csme/meas/papers/wiliam.html

Winograd, T., & Flores, F. (1986). *Understanding computers and cognition*. Norwood: Ablex.

Winsor, D.A. (1999). Genre and activity systems: The role of documentation in maintaining and changing engineering activity systems. *Written Communication, 16*, 200-224.

Yates, J., & Orlikowski, W. (1992). Genres of organizational communication: A structurational approach to studying communication and media. *Academy of Management Review, 17*, 299-326.

Yeo, G. (2007). Concepts of record (1): Evidence, information, and persistent representations. *American Archivist, 70*, 315-343.

Yeo, G. (2008). Concepts of record (2): Prototypes and boundary objects. *American Archivist, 71*, 118-143.

Yeo, G. (2010). Representing the act: Records and speech act theory. *Journal of the Society of Archivists, 31*, 95-117.

Yeo, G. (2015). Record(s). In L. Duranti & P.C. Franks (Eds.), *Encyclopedia of Archival Science*. Lanham: Rowman & Littlefield.

Yoshioka, T., Herman, G., Yates, J., & Orlikowski, W. (2001). Genre taxonomy: A knowledge repository of communicative actions. *ACM Transactions on Information Systems, 19*, 431-456.

ARNOUD GLAUDEMANS AND JACCO VERBURGT

The Archival Transition from Analogue to Digital: Revisiting Derrida and Flusser

I. Introduction

1. Any account of present-day archives should not only address practical, operational or managerial issues but also explicate the relevant theoretical issues regarding the specific nature and societal impact of digital information – if only because practical, operational or managerial issues, important as they obviously are, always presuppose some underlying theoretical framework. Unfortunately, such underlying views often remain somewhat implicit in current debates on digital archives. In this article, we aim to meet this 'lack of explicitness' by discussing – that is, exploratively comparing and contrasting – two important theoretical views concerning digital information, notably in order to unravel possible theoretical 'blind spots' *and* to deal with practical issues more adequately, for instance regarding the (re-) use of digital archives and governmental accountability.

    In section II, we address Jacques Derrida's (so-called deconstructivist or postmodernist) views on the archive in his well-known text 'Mal d'archive' (Derrida, 1995 [1994]).[1] In this text, Derrida provides us with an analysis of the notion 'archive'. His point of departure is a distinction between 'mneme' or 'anamnesis' (living memory or recollection) on the one hand, and 'hypomnema' (records or public records) on the other. He states for instance: "Let us never forget this Greek distinction between *mněmě* or *anamněsis* on the one hand, and *hypomnema* on the other. The archive is hypomnesic." (Derrida 1995 [1994], p. 14) And he suggests that, contrary to classical (metaphysical and scientific) thought, this hypomnesic character of the archive is fundamental and irreducible, that is to say, "the idea of an archive properly speaking, of a hypomnesic or technical archive, [...] cannot be reduced to memory: neither to memory as conscious reserve, nor to memory as rememoration" (Derrida 1995 [1994], p. 58). According to Derrida, this irreducibility ultimately confronts us with a rather troubled situation, which he calls 'mal d'archive': "never to rest, interminably, from searching for the archive right where it slips away" (Derrida 1995 [1994], p. 57). In our view, his approach embodies a highly acclaimed

---

[1] Derrida's lecture text, initially entitled 'The Concept of the Archive: A Freudian Impression', will be quoted according to the first English translation from 1995 ('Archive Fever: A Freudian Impression', *Diacritics*, Vol. 25, No. 2, pp. 9-63). We will not use the 1996 translation. See Literature.

and widely adopted stance on the archive, especially on the important notion of scriptural recording. However, an explicit account of the concrete, that is functional and technical, aspects of digital information is largely missing.

In section III, we address Vilém Flusser's less well-known analysis of what he calls the technical image – a concept which includes the digital – in his intriguing book *Ins Universum der technischen Bilder* (Flusser, 1996 [1985]).² In this text, one of the major concerns and objectives is to explicate, and account for, a distinction between 'imaginieren' (belonging to the realm of what he calls traditional images) and 'einbilden' (belonging to the realm of technical images). For example, he writes: "Traditional images are mirrors. They capture the vectors of meaning that move from the world toward us, code them differently, and reflect them, recoded in this way, on a surface. Therefore, it is correct to ask what [*was*] they mean. Technical images are projections. They capture meaningless signs that come to us from the world (photons, electrons) and code them to give them a meaning. So, it is incorrect to ask what they mean (unless one gave the meaningless answer: they mean photons). With them the question to ask is, what is the purpose [*wozu*] of making the things they show mean what they do? For what they show is merely a function of their purpose." (Flusser 2011 [1985], p. 48) In our view, this crucial distinction implies a constructivist, functionalist, and modernist approach to digital information and, most importantly, enables to understand the digital archive in a more adequate, techno-functional or content-technical, way.

In section IV, we systematically reinterpret Derrida's conception and questions regarding the archive in light of Flusser's conceptual framework and distinctions. We readdress the essential elements of Derrida's conception of the archive, which we believe are in need of reinterpretation, in light of Flusser's account of, and emphasis on, the technical aspects of the digital.

Finally, in section V, the conclusion, we not only briefly resume our findings in sections II-IV, but also formulate five more or less programmatic questions and theses concerning the archive and the digital.

2. Before turning to section II, let us first illustrate the difference in approach between Derrida and Flusser by quoting their remarkable descriptions of the apparently trivial experience of pressing keys on a keyboard (of a typewriter or a computer). First a quotation from Derrida's text: "[..] while tinkling away on my computer. I asked myself what is the moment proper to the archive, if there is such a thing, the instant of archivization strictly speaking, which is not, and I will come back to this, so-called live or spontaneous memory (*mněmě* or *anamněsis*), but rather a certain hypomnesic and prosthetic experience of the technical substrate. Was it not at this very instant that, having written something or other on the screen, the letters remaining as if suspended and floating yet at the surface of a liquid element, I pushed a certain key to "save" a text undamaged, in a hard and lasting way, to protect marks from being erased, so as thus to ensure salvation and *indemnity*, to stock, to accumulate, and, in what is at once the same thing and something else, to make the sentence thus available for printing and for reprinting, for reproduction?" (Derrida 1995 [1994], p. 22) In this quote, a particular focus on the save key is discernible.

By doing so, Derrida prioritises the hypomnesic and recording character of archivisation. The latter notion is of course well known in archival science. In section II of this article, we stress the importance of the formal aspects of the hypomnesic in Derrida's text (namely sign, materiality, place, and repetitivity).[3]

In the following passage from Flusser's book, the key use seems to be limited (not unlike Derrida) to the linear recording, or 'saving', of a text. In the second part of the quote, however, Flusser clearly prioritises the computational aspect of key use and links it – albeit retrospectively – to the emergence of technical images. Here is the passage: "As I run my fingertips selectively over the keyboard of my typewriter to write this text, I achieve a miracle. I break my thoughts up into words, words into letters, and then select the keys that correspond to these letters. I calculate [*'kalkuliere'*] my ideas. And the letters then appear on the piece of paper that has been put into the typewriter, each for itself, clear and distinct, and nevertheless forming a linear text. The typewriter computes [*'komputiert'*] what I have calculated. It succeeds in packaging the particles into rows. That is a miracle, despite the transparency of the process. [..] By observing how images are synthesized on a computer screen by pressing keys, we can, looking back in a sense, recognize the miracle of mechanical button pressing as well: it is the miracle of calculation followed by computation, the miracles to which technical images owe their existence." (Flusser 2011 [1985], p. 24) From this quote, it is clear that the emergence of technical images (Flusser's key concept) is based on calculations and, especially, computations – the digital being the perfect form of the technical image through the technique of bit encoding (0/1), which in turn enables algorithmic processing. Therefore, Flusser stresses that we should emancipate ourselves from being merely "[b]utton-pressing functionaries" (*tastendrückenden Funktionäre*), only able of using save, send, and receive keys, and adopt what he calls "the computing touch" (*komputierendes Tasten*). For our telematic and dialogical society, as Flusser calls and envisions it, demands or requires a more conscious use of keys.

II. Revisiting Derrida: relevant concepts and pertinent questions

3. In this section, we further explore Derrida's theoretical views on the archive – his archival theory, if you will – by discussing a few relevant concepts or conceptual distinctions he uses and a few pertinent questions and issues he raises. Obviously, it is impossible to discuss all the ins and outs of Derrida's quite complicated and layered text, let alone his wider oeuvre, which – incidentally – still seems to be highly controversial, both among philosophers and archival theorists and scientists.

---

[2] Flusser's book was first published in 1985 (edited by Andreas Müller-Pohle). In this article, we are referring to the English translation from 2011. Whenever necessary, we have quoted Flusser's original German terminology from the fifth edition (1996). See Literature.

[3] In our view, much of the archival literature on Derrida aims to account for archivisation, not primarily in terms of these formal aspects of the archive, but rather in terms of social and cultural contexts (or factors). This socio-cultural or social constructivist approach seems to involve what Ketelaar calls 'archivalisation': "Before *archivization* [...] is another 'moment of truth'. It is *archivalization*, a neologism which I invented, meaning *the conscious or unconscious choice (determined by social and cultural factors) to consider something worth archiving*." (Ketelaar, 2001, p. 133) In this article, we do not address the question as to whether such a perspective on archivisation is sufficiently adequate to account for the hypomnesic characteristics Derrida ascribes to the archive.

For present purposes, it suffices to discuss the following: Firstly, we briefly outline Derrida's general views concerning the archive, also in relation to the well-known notion of archivisation; secondly, we discuss three basic elements of his conception of the archive: mediality, domiciliation or topo-nomology, and consignation; and finally, his suggestion of an entirely different logic at work in the digital context, which he raises only in the margins of his text, but which might be fruitfully discussed in terms of Flusser's theoretical framework, as we argue in section III.

4. As to Derrida's general views, the first thing to note is that he uses two key terms: 'archontic' and 'archivisation'. The first term he derives from an etymological analysis of the Greek term 'arkhe' or 'arkheion', an important term which we address shortly (see #6). By the second term he aims to designate a concrete act of archiving or recording, for instance the save-key example from the introduction. Moreover, what he aims to designate by 'archivisation' is not so much a general concept of recording acts, but rather an always particular case of recording, which determines the actual meaning of a record in its specific context. Another important thing to note is Derrida's emphasis on the irreducible, external and hypomnesic character of the archive:

> [...] the archive, if this word or this figure can be stabilized so as to take on a signification, will never be either memory or anamnesis as spontaneous, alive and internal experience. On the contrary: the archive takes place at the place of originary and structural breakdown of the said memory. There is no archive without a place of consignation, without a technique of repetition, and without a certain exteriority. No archive without outside. (Derrida 1995 [1994], p.14)

5. This quote clearly indicates that the archive is always externally materialised and different from our inner experienced live memory. It does not simply present or represent our live memory. Rather, in order to function, the archive has to be something exterior to, and apart from, live memory. It needs to consist of certain signs, to have a certain materiality, to be at a certain place, and to involve a certain technique of repetition. Unlike Derrida, one could use the term *mediality* to cover these aspects of exteriority. In our view, abstracting from mediality would result in mistaking a particular figure of the archive for a general concept of it. Hence, having a specific technical substrate – e.g., ink on paper as to the alphabet, microfilm as to images, some physical (analog) carrier as to digital information – with its own techno-functional (im)possibilities is presupposed by, or intrinsic to, any archive.

6. At the very beginning of his text, Derrida introduces the second element of his conception of the archive, namely the crucial element of domiciliation or topo-nomology, by means of an etymological analysis of the Greek term 'arkheion'. The full passage is quoted here:

> As is the case for the Latin *archivum* or *archium* [...], the meaning of "archive," its only meaning, comes to it from the Greek *arkheion*: initially a house, a domicile, an address, the residence of the superior magistrates, the

archons, those who commanded. The citizens who thus held and signified political power were considered to possess the right to make or to represent the law. On account of their publicly recognized authority, it is at their home, in that *place* which is their house (private house, family house, or employee's house), that official documents are filed. The archons are first of all the documents' guardians. They do not only ensure the physical security of what is deposited and of the substrate. They are also accorded the hermeneutic right and competence. They have the power to interpret the archives. Entrusted to such archons, these documents in effect state the law: they recall the law and call on or impose the law. To be guarded thus, in the jurisdiction of this *stating the law*, they needed at once a guardian and a localization. Even in their guardianship or their hermeneutic tradition, the archives could neither do without substrate nor without residence. (Derrida, 1995 [1994], pp. 9-10).

From this quote it is clear that the term 'archive' designates both a collection of documents and the building in which they are kept. There can be no archive without domiciliation. For a collection of documents to be an archive, there has to be a certain privileged place where it resides; it has to be under both physical and interpretative control and guardianship; for only then this collection of documents can count as stating the law. In other words, documents are kept and classified under the title of an archive by virtue of a privileged topology and nomology.

7. The third element of Derrida's conception of the archive, namely the function or power of consignation, must be understood as the act of carrying out the second element; it is its executive counterpart. However, Derrida stresses that consignation does not only concern acts of recording, or the results thereof, but also what any act of recording (*consignatio*) always already presupposes:

> By consignation, we do not only mean, in the ordinary sense of the word, the act of assigning residence or of entrusting so as to put into reserve (to consign, to deposit), in a place and on a substrate, but here the act of consigning through *gathering together signs.* It is not only the traditional *consignatio*, that is, the written proof, but what all *consignatio* begins by presupposing. *Consignation* aims to coordinate a single corpus, in a system or a synchrony in which all the elements articulate the unity of an ideal configuration. In an archive, there should not be any absolute dissociation, any heterogeneity or *secret* which could separate (*secernere*), or partition, in an absolute manner. The archontic principle of the archive is also a principle of consignation, that is, of gathering together. (Derrida 1995 [1994], p. 10)

It is fair to say that Derrida's phrase "what all *consignatio* begins by presupposing" equals what is commonly known as the (defining) context of the archive. And the context of an archive, any particular archive, is defined or determined by questions such as: Which types of acts are to be recorded? Which aspects of these acts are to be recorded? How are these to be classified and incorporated into the archive? And according to which criteria or rules?

Of course, many more (detailed) questions could be added here. Our aim, however, is not to be comprehensive, but simply to underline the double meaning of Derrida's notion of consignation, denoting both result and presupposition.

8. But what, then, would be the implications of Derrida's threefold conception of the archive, especially regarding the digital? First of all, recall Derrida's emphasis on the hypomnesic nature of the archive, that is, his account of the exteriority – or, in our terms, mediality – of the archive. Consider the following quote:

    > […] the archive, as printing, writing, prosthesis, or hypomnesic technique in general is not only the place for stocking and for conserving an archivable content *of the past* which would exist in any case, such as, without the archive, one still believes it was or will have been. No, the technical structure of the *archiving archive* also determines the structure of the archivable content even in its very coming into existence and in its relationship to the future. The archivization produces as much as it records the event. This is also our political experience of the so-called news media. (Derrida 1995 [1994], p. 17)

    In this passage, interestingly enough, Derrida seems to go so far as to claim that the hypomnesic even determines the very structure of the archivable content, rather than being just a replaceable recording technique or a presupposed context of already existing records. In addition, this claim seems to involve a fundamentally different way of being affected by the digital. And, indeed, Derrida seems to be quite aware of such more experiential implications:

    > Is the psychic apparatus *better represented* or is it *affected differently* by all the technical mechanisms for archivization and for reproduction, for prostheses of so-called live memory, for simulacrums of living things […] (microcomputing, electronization, computerization, etc.)? (Derrida 1995 [1994], p. 16).

    This question seems to be a rhetorical one, which would imply that, according to Derrida, we are indeed affected very differently by the digital. And, in fact, it is not hard to find more and ample textual evidence to the support and confirm the thesis, our thesis, that Derrida holds this view. Consider the following quote, in which he compares the present-day digital archive with the situation after an earthquake and confesses that he would have liked to talk about the huge impact of (late twentieth century) digital technologies on the traditional or classical (late ninetieth, early twentieth century) archive in terms of what he calls retrospective science fiction:

    > I would have liked to devote my whole lecture to this retrospective science fiction. I would have liked to imagine with you the scene of that other archive [the digital archive] after the earthquake and after the "*après-coups*" of its aftershocks. This is indeed where we are. (Derrida 1995 [1994], p. 17).

From a present-day perspective, from "where we are" today, it seems obvious and rather trivial to say that we are affected by new digital developments. But what, exactly, does it mean to say that we are really affected differently? Unfortunately, Derrida does not really go into detail here. The only example he does refer to is e-mail, by stating that electronic mail is "on the way to transforming the entire public and private space of humanity, and first of all the limit between the private, the secret (private or public), and the public or the phenomenal" and that it is "not only a technique" and "must inevitably be accompanied by juridical and thus political transformations", including "property rights, publishing and reproduction rights" (Derrida 1995 [1994], pp. 17-18). One could of course sum up a lot of other examples apart from e-mail, such as big data, dynamic simulating models, and social media. But the crucial point, Derrida's crucial point, should not be missed: namely that we are dealing with (or are affected by) an entirely different logic, a fundamentally different spatial and temporal logic, which opposes and no longer fits the (traditional) logic of continuous representation. Because, as Derrida puts it, "if the upheavals in progress affected the very structures of the psychic apparatus, for example in their spatial architecture and in their economy of speed, in their processing of spacing and of temporalization, it would no longer be a question of simple continuous progress in representation, in the *representative* value of the model, but rather of an entirely different logic." (Derrida 1995 [1994], p. 16)

III. Revisiting Flusser: conceptual framework and distinctions

9. The interesting issue Derrida raises in the previous section is, in our view, his reference to an entirely different, non-representational or non-representative, logic at work in the digital context. What would such a logic involve and imply? Not only regarding what Derrida refers to as the archontic function (or principle, power etc.) of the archive as we know it, but also taking into account that digital machineries do not imitate (a pre-given) reality, as Slavoj Zizek rightly signals, but rather generate and simulate semblances of a non-existing reality – so that information in fact would consist of simulacra (see # 11)?

    In order to deal with these questions, it is helpful to read Flusser's analysis in terms of Derrida's notion of retrospective science fiction, the subject matter of this science fiction being "the scene of that other archive after the earthquake and after the "*après-coups*" of its aftershocks", and "other archive" meaning the digital archive, and stressing that this is indeed "where we are" (see #8). In our view, Flusser's theory of the technical image not simply addresses the issue of an entirely different logic, but also – unlike Derrida – provides a theoretical framework to account for the specific content-technical characteristics of digital media and information in a more substantial way. In addition, Flusser motivates his focus on the apparently accidental or arbitrary means and techniques of dealing with (digital) information by arguing that these have a decisive influence on our lives: "the structure through which information is carried [*Informationsträger*] exerts a decisive influence on our lives [*Lebensform*]." (Flusser 2011 [1985], p. 5)

10. Let's start by giving a short outline of Flusser's general theoretical framework. Basically, Flusser's general theoretical framework consists of five layers or rungs (*Stufen*) of abstracting (*abstrahieren*). In the English translation, these layers are described as follows:

- *First rung*: Animals and "primitive" people are immersed in an animate world, a four-dimensional space-time continuum of animals and primitive peoples. It is the level of concrete experience.
- *Second rung:* The kinds of human beings that preceded us (approximately two million to forty thousand years ago) stood as subjects facing an objective situation, a three-dimensional situation comprising graspable objects. This is the level of grasping and shaping, characterized by objects such as stone blades and carved figures.
- *Third rung: Homo sapiens sapiens* slipped into an imaginary, two-dimensional mediation zone between itself and its environment. This is the level of observation and imagining characterized by traditional pictures such as cave paintings.
- *Fourth rung*: About four thousand years ago, another mediation zone, that of linear texts, was introduced between human beings and their images, a zone to which human beings henceforth owe most of their insights. This is the level of understanding and explanation, the historical level. Linear texts, such as Homer and the Bible, are at this level.
- *Fifth rung*: Texts have recently shown themselves to be inaccessible. They don't permit any further pictorial mediation. They have become unclear. They collapse into particles that must be gathered up. This is the level of calculation and computation, the level of technical images. (Flusser 2011 [1985], pp. 6-7)

For present purposes, in this section, we focus on the transition from layer four to layer five by discussing three conceptual distinctions, which are partly based on, and partly inferred from, Flusser's text. Firstly, the distinction between the representational and the simulative; secondly, distinction between to imagine (*imaginieren*) and to envision (*einbilden*), as already mentioned in the introduction; and thirdly, the distinction between the discursive or linear and the dialogical.

11. As to the first distinction, it is inducible from Flusser's theory – even though he does not literally employ this terminology – that the digital is not so much ruled by a *representational* logic but rather by what one could call a *simulative* logic. This difference between representation and simulation is intrinsically linked to the both technical and conceptual difference between analog and digital, between inscription and encoding. Essentially, the difference between analog and digital, between inscription and encoding, lies in the use of the physical. Analog techniques use physical, in itself analog, phenomena for the sake of recording and carrying only a *specific kind* of information. Sometimes a machine is needed (e.g., a microfilm) to make analog information sensorially available, sometimes there is not (ink on paper). Digital techniques, by contrast, use these in itself analog – continuous, non-discrete – phenomena for encoding and transferring *any kind* of information in discrete bits (yes/no or 0/1, and nothing

in between), which Flusser refers to as "Punktelemente". To make this digital information graspable, however, we also still need analog physical phenomena, such as screens and speakers. For bits and algorithmic processes as such are not sensory.

Here, again, one could raise the question whether we are actually 'affected differently', as Derrida suggests (see #8) by digital information, or is it merely 'better represented' to us, in comparison to analog media? One could also recall McLuhan's observation that the message of a newly introduced medium (in this case the digital) tends to be the previous medium (in this case the analog). The first use of the new one is to imitate the old (cf. McLuhan 1994, pp. 7-21). What makes the difference here is, firstly, that the basic technique of bit-encoding makes information extremely flexible, processable, and transportable (e.g., the internet). When compared to analog information, there are almost no physical constraints. Secondly, because specific kinds of information are not bound to, or hard-coded on, specific physical phenomena (e.g., music as magnetic fields in the case of a cassette recorder), the encoding technique also does not restrain on form and structure of information. Indeed, a lot of today's digital information has no analog counterpart (e.g., hyperlink, database etc.). Although bits undoubtedly have a physical side (photons, electrons, etc.) we are dealing with an almost immaterial technical substrate. In fact, Flusser describes the digital as the perfect realm of the technical image, because the restraints on form and structure of technical images are reduced to an absolute minimum. In our view, continuous physical representation, for instance on a two-dimensional screen, is *not* the basic way of making sense of the principally non-sensory and algorithmically processed bits, including all the possible, analogically unprecedented, forms and structures of information. Therefore, we are indeed affected by, and oriented towards, something *beyond* the screen that, as such, is not physically present, nor physically representable, but nonetheless perfectly real, albeit in a non-empirical and non-empiricist way, which might be linked to the German word 'wirklich'. As Slavoj Zizek puts it:

> VR [Virtual Reality] doesn't imitate reality, it simulates it by way of generating its semblance. Imitation imitates a pre-existing real-life model, whereas simulation generates the semblance of a non-existing reality – it simulates something that doesn't exist. (Zizek, 1997)

This conception of the simulative nature of digital information neatly corresponds to Flusser's idea that technical images do not represent something in the world 'out there', but project a meaning, some informational content, without designating or mirroring something outside of it (see #1). The basic form of digital information is therefore that of simulation. In fact, digital information consists of simulacra. Or, in other words, the digital is ruled by a simulative logic.

12. The second distinction, the one between imagine (*imaginieren*) and envision (*einbilden*), corresponds to our first distinction between representation and simulation. Within Flusser's general framework, the concepts of imagine and envision should be located on the third/fourth and the fifth layer, respectively. Ultimately, as to the fourth and fifth layer, his framework is

[…] meant to show that technical images and traditional images arise from completely different kinds of distancing from concrete experience. It is meant to show that technical images are completely new media, even if they are in many respects reminiscent of traditional images. They "mean" in a completely different way from traditional images. (Flusser 2011 [1985], p. 7)

But what, exactly, does the difference between traditional and technical images consist of? According to Flusser, traditional images aim to represent or mirror a pre-given reality, thus generating meaning (*was*), while technical images aim to project or simulate images on the basis of atomic or elementary particles (*Punktelemente*), thus generating purely heuristic, or cybernetic, functionality (*wozu*). Recall the quote from the introduction:

Traditional images are mirrors. They capture the vectors of meaning that move from the world toward us, code them differently, and reflect them, recoded in this way, on a surface. Therefore it is correct to ask what [*was*] they mean. Technical images are projections. They capture meaningless signs that come to us from the world (photons, electrons) and code them to give them a meaning. So it is incorrect to ask what they mean (unless one gave the meaningless answer: they mean photons). With them the question to ask is, what is the purpose [*wozu*] of making the things they show mean what they do? For what they show is merely a function of their purpose. (Flusser 2011 [1985], p. 48)

A defining difference of traditional and technical images is the difference in dimensionality. Traditional images, which originate at the third rung of abstracting in Flusser's framework, are two- or three-dimensional physical objects used to mirror and thus represent a meaning. On the other hand, technical images, which only originate at the fifth rung of abstracting, consist of zero-dimensional or dimensionless simple particles, out of which projections or simulacra are constructed by means of technical processing. This difference in dimensionality precisely defines the completely different nature of 'imaginieren' and 'einbilden'.

13. The distinction between discursive or linear and dialogical, the third and last distinction we discuss in this section, is illustrated and explained by Flusser through a description of the nature of key usage on machines:

[…] it is in the character of keys to link up with one another "in dialogue" (e.g., through cables) to form networks, that is, to operate not as discursive but rather as dialogical instruments [*Vorrichtungen*]. The difference between sending and receiving, between productive and reproductive keys, is therefore to be viewed as provisional. The typewriter is only a forerunner of the telewriter, the control panel of the washing machine only a forerunner of a feedback loop linking manufacturers and the users of washing machines. And the current state of keys in general is only a forerunner of a telematic society. (Flusser 2011 [1985], p. 30)

Again, the defining difference between the discursive and the dialogical is one of dimensionality. The discursive or linear, originating at the fourth layer of abstracting in Flusser's framework, is one-dimensional, for instance scriptural and linear texts, which contain concepts and process them in order to represent and exchange meaning. As to the dialogical, projections or simulacra are algorithmically processed out of zero-dimensional or dimensionless simple particles, as well as directly linked and made (ex)changeable through keys. Thus, this difference in dimensionality also defines the different nature of the discursive and the dialogical. In addition, at the dialogical level, there is a feedback-loop of technical images and individual persons, which is different in comparison to the discursive:

> A feedback loop must appear between the image and the receiver, making the images fatter and fatter. The images have feedback channels that run in the opposite direction from the distribution channels and that inform the senders about receivers' reactions, channels like market research, demography, and political elections. This feedback enables the images to change, to become better and better, and more like the receivers want them to be; that is, the images become more and more like the receivers want them to be so that the receivers can become more and more like the images want them to be. That is the interaction between image and person, in brief. (Flusser 2011 [1985], pp. 53-54)

A first after-effect of the technique of dialogical linkage is that it replaces scriptural, discursive, texts. This implies at least a relativisation of the historical, time-as-linear, mode of understanding and criticising. From the zero-dimensional or dimensionless dialogical perspective, however, time is not pre-given as a (one-dimensional, historical) line from past to future. In this sense, the dialogical is strictly post-historical. This implies that, in the realm of telematic society, historical 'Kulturkritik' becomes obsolete and another method of critique or criticism has to be developed. Such a new form of critique should take into account that technical images do not mean to represent, so that it would be a mistake to criticise *what* they represent:

> [...] another look at the possibilities that lie dormant in telematic equipment [*Vorrichtungen*], at the silly twiddling with telematic gadgets, shows where most cultural critics go wrong. They try to criticize the radiating centers to change or do away with them. But revolutionary engagement has to begin not with the centers but with the silly telematic gadgets. It is these that must be changed and changed in ways that suit their technology. Should this be successful, the centers will collapse of their own accord. No longer historical but rather cybernetic categories must be used for criticism. (Flusser 2011 [1985], p. 86)

A second after-effect of the technique of dialogical linkage – or aftershock as Derrida calls it and preludes to via the example of e-mail (see #8) – is the disappearance of the distinction of, on the one hand, the private and, on the other, the public and political space. The discursive or linear mode of communication presumes a send-receive-save usage of keys: to send information

from my private space into the public space, to be received by others in their own private space. The essential difference of the discursive and the dialogical, as described by Flusser, is that the latter presumes a live, two-way, feedback by means of direct linkage in the *same* space. This is the case in our daily use of digital media (e.g., click, swipe, etc.). The fact that we are sender and receiver at the same time, in the same space, implies that the discursive distinction of public and private tends to disappear: "[...] keys have burst the boundaries between private and public. They have blended political with private space and made all inherited conceptions of discourse ['*Diskurs'*] superfluous." (Flusser 2011 [1985], p. 30)

IV. A Reinterpretation of Derrida in Light of Flusser

14. We will now turn to a reinterpretation of Derrida's conception of the archive by means of the concepts and distinctions described in the previous section, which aimed to outline Flusser's account of the specifics of digital mediality and its main characteristics, as well as its differences from analog mediality. In our view, it is to be expected that these differences have a major impact on nature, function, and position of the archive. In this section, we will first address both the topological and nomological aspects of the conception of the archive. Subsequently, we consider the aspects of domiciliation, since – as Derrida rightly stresses – for a collection of documents to effectively be an archive they have to be held at a certain 'privileged place', in which the topological and nomological intertwine. We close this section with a reinterpretation of Derrida's concept of consignation.

The archontic principle presupposes physical guardianship and control of the material, the technical substrate, of the archive. As to the locational or topological aspect of the digital, it follows from our account of digital mediality (see #11) that we are *not* primarily dealing with a physical place or topology but rather a simulated, 'virtual', place and topology generated by means of algorithmic processing and usually mediated through a computer screen. Of course, bits and their algorithmic processing also have a physical (mainly electrical) side, but it is impossible and irrelevant to map the (digital) technical image to a physical location – particularly given present-day virtualisation and cloud technology. Note, also, that even a docx – in a file explorer window – has nothing physical, bit-like, about it.

Therefore, guardianship and control of the digital archive should concern itself with 'virtual' place and topology. Apart from exceptional situations (such as the now withdrawn Patriot act), physical guardianship and control of – more and more worldwide distributed – bits cease to be a relevant part of the figure of the digital archive. In addition, the technical substrate to guard and control, namely bits, should be understood as *effectively* immaterial, given their radical transferability and flexibility. In this vein, governmental guardianship and control over digital archives should concern itself not only with document-like records but also, and primarily, with managing technical and informational architectures or models ('topologies') and with safeguarding the transparency of algorithmic processing, e.g. through open source licensing.

15. As to the nomological aspects of Derrida's conception of the archive, that is the embodiment and objectification of the law (see #6), Flusser's analysis indicates, firstly, that such a law is to be located in the realm of technical images. Secondly, it also indicates that the actual forms and structures of technical images objectifying the law may greatly vary, and differ from document-like records (e.g., a website). Thus, a government that wishes to prescribe, interpret, and execute the law explicitly and consciously, needs to actively concern itself with technical images or simulacra objectifying the law, by actively managing its topology and safeguarding the transparency of algorithmic processing. Otherwise it will become increasingly unclear *what* and *where* the law actually is. The law will become increasingly opaque, non-transparent, and blurry – not only with regard to civic society, but also within governmental institutions themselves. The same concern applies to establishing *who* is to be held accountable, especially when it comes to executing intergovernmental and privately outsourced governmental tasks in the same topological space (e.g., cooperative digital environments). One could interpret intergovernmental cooperation in the chain- or network-wise execution of tasks as a 'dialogification' in a Flusserian sense and the blurring of the distinction between public and private as an after-effect of this dialogification (see #13).

In fact, this account of the nomological situation implies that it is counterproductive, and possibly undesirable from the viewpoint of democratic accountability, to prioritise a document-like concept of a record when dealing with the digital. Since the law is no longer reducible to, and fully conceivable as, a document-like record – and increasingly so. In our view, this conclusion is not fundamentally affected by complications resulting from the still often combined (hybrid) usage of both paper and digital media for archiving. A 'conservative' stance on digitalisation might have undesirable effects from the viewpoint of democratic accountability: the nomological *what, where*, and *who* threatens to become blurry and loses its necessary transparency.

16. As to domiciliation or topo-nomology, perhaps the most important element of Derrida's conception of the archive, let us first recall that he defines it as the privileged place of the archive, namely the place where the physical control and guardianship happens ('topology') *and* from which the prescription, interpretation, and execution of the law takes place ('nomology'). From our analysis (see #14-15) a *different* basic form of topo-nomology, inherent to the digital, seems to arise. The basic structure or traditional model of domiciliation, involving an 'archontic center', from which things are hierarchically ruled and managed, seems to have broken into pieces, if not disappeared at all.

Flusser argues that in a utopian-telematic society the basic archontic centralistic form of ruling and governing will become dysfunctional and will be replaced by the topo-nomological form inherent to the cybernetic techniques by means of which ruling and governing will actually take place:

> In the universe of technical, telematic images, there is no place for authors or authorities. Both have become superfluous through the automation of production, reproduction, distribution, and judgment. In this universe,

images will govern the experience, behavior, desire, and perceptions of individuals and society, which raises the question, what does govern mean when no decisions need to be made and where administration is automatic? In a telematic society, does it still make sense to speak of government, of power and the powerful? [...] If politics is understood to be the art of informing, then the question becomes how rather than what: in a telematic society, how does governing, the exercise of power, the administration of justice occur? To go straight to the obvious answer, cybernetically. I am defining *cybernetic* here – without claiming general applicability – as automatic guidance and control of complex systems to take advantage of improbable accidents and to generate information. (Flusser 2011 [1985], pp. 123 and 125)

As this passage indicates, the basic model or structure of governing and ruling, in Flusser's telematic society, is not that of a central brain that rules, decides, commands, etc., but rather a model which resembles an ant heap or formicary consisting of 'networked' individual brains: "According to this model, the telematic society is a structure in which human brains follow the same cybernetic methods as ant brains. They function for one another, and function predominates." (p. 130) However, Flusser adds an important adjustment to his metaphor of an ant heap or formicary consisting of networked individual brains, namely that the network structure of telematic society has no 'outside' to relate to. In this sense, telematic society entails a double reversal of the traditional archontic model. Firstly, the primary center is distributed across, essentially equal, nodes or junctions of a network – up to the point that there is no primary center anymore (topology). Secondly, the clock speed of the linear rhythm of sending and receiving, from the center and back, is increased by direct, live, dialogical linkage of the aforementioned nodes and junctions – up to the point that the linear becomes a point (nomology). And one could wonder whether the Derridian notion of topo-nomology is still useful, functional or applicable here, because digital domiciliation would include the whole of telematic society. The notion of domiciliation tends to become dysfunctional in the sense that it ceases to deliver concepts and distinctions by means of which we can understand archives and society.

17. The last element of Derrida's conception of the archive, which we aim to reinterpret in this section, is the element of consignation. Let's first draw some conclusions from our analysis so far. The disappearance of the archontic principle in Flusser's conception of a utopian-telematic society (see #16) seems to eliminate the very *raison d'être* of the archive as we know it, namely the accommodation and safeguarding of governmental accountability and the subsequent (re-)use for civic, scientific (e.g., historical), or perhaps even entrepreneurial, purposes. Although a lot of current developments in the archival field can be interpreted in terms of telematic society announcing and enforcing itself, it would be premature to conclude that the archive and its functions are simply and totally becoming obsolete.

As explained above (see #7), Derrida's conception of the function of consignation must be understood as the executive counterpart of, and within, any particular topo-nomological setting. Additionally, Derrida stresses

that consignation does not only concern the executive part (including the results thereof), but also what this part always already presupposes. Given the current transitional situation of, on the one hand, telematic society and its intrinsic structure of networked nodes enforcing itself, and, on the other hand, governments and their intrinsic archontic structure still existing, one could conclude that there are actually *two* different areas or settings where consignation takes place – and that both need to be addressed theoretically and practically. The first, more traditional, area would be that of, primarily, accommodation and safeguarding of governmental accountability, in which the guardianship and control of digital information is the crucial topic. The second, newer, area would be that of the accommodation and safeguarding transparency of algorithmic processing and the subsequent (re-)usability for dialogical purposes.

V. Conclusion

18. Both Derrida and Flusser provide highly valuable insights into the present day archival transition from analog to digital, or from scriptural and discursive records to algorithmically processed and dialogical technical images. In our view, Derrida scrutinises the traditional function and positioning of an archive as determined by the concept of a record, as well as its specific physical and institutional context. Although he is very much aware of the possible impact of technology, including digital technologies, he fails to sufficiently thematise important content-technical aspects of the present day digital archive. In a way, Derrida remains captivated by the archival promise of delivering live memory. Flusser, on the other hand, focuses precisely on the concrete content-technical nature of the digital. His analysis aims to envision a utopian, 'science-fictional', cybernetically governed society, in which the historical or commemorative function of an archive will ultimately become dysfunctional and obsolete. However, we argue that democratic accountability and control still presupposes a traditional conception of the archive. Hence, it is imperative that the appropriate recording functionality should be 'built in' to the digital.

19. To close this article, and in view of further discussion, we formulate five more or less programmatic questions, or theses, regarding the archival transition to the digital.

    i.   As to archival science or theory: Should the record-based approach to the archive be replaced by the functional approach in terms of algorithmic processing? Obviously, all digital information, including digital records, is processed algorithmically. Our thesis is that the impact on the archival function of the intrinsic, technical, effects of digital mediality (without analog counterpart), can be better accounted for from the functional approach.

    ii. As to data science: Could existing archival concepts, and the specific detailed archival knowledge they entail, be (re-)used in the approach of algorithmic processing, to enhance and enrich the desired functionality in this particular field? And which particular concepts would be relevant in this respect? One could think, here, of concepts like authenticity, provenance, and accessibility.

iii. As to cultural criticism ('Kulturkritik'): Which function(s) should the archive and archival theory fulfill today, especially in light of Flusser's distinction between discursive and dialogical (see #13)? It seems that historico-cultural criticism, in which the archive functions as a source, is becoming increasingly obsolete and even counterproductive. Should the comparative approach of archival science (cf. Ketelaar, 1997) not be complemented by an account of the archival dialogical function within a critique of telematic society (e.g., open data)?

iv. As to governmental accountability and control: Should the effects of the current and progressing transition to the digital not be accounted for better, in order to avoid an increasingly diffuse allocation of responsibility? In our view, a lacking account of the impact of digitalisation will have undesirable effects on democratic accountability (see #15-16).

v. As to specific archival recording or processing machinery: Given the rapidly evolving functionalities of standard cloud technology, what would (still) be necessary or desirable, particularly in order to meet commonly shared criteria of availability and accessibility, in the context of a telematic society? How could, for instance, a dialogical (re-)use, including a commemorative (re-)use, be accommodated and safeguarded?

Of course, these questions could be complemented, amended or specified further. Our main aim in this article was to detect possible theoretical 'blind spots', notably for the sake of dealing with urgent practical issues more adequately.

Literature

Derrida, Jacques (1995 [1994]). Archive Fever: A Freudian Impression. *Diacritics, 25*(2), 9-63. (Translated by Eric Prenowitz; Derrida's original lecture, initially entitled 'The Concept of the Archive: A Freudian Impression', was given on June 5, 1994, during a London colloqium on *Memory: The Question of Archives*; and the later modified, now commonly used, French title of the text is 'Mal d'archive: Une impression freudienne', or simply 'Mal d'archive', often referring to the separate edition at Éditions Galilée (Paris 1995; reedited in 2008); note that we do not refer to the separate English edition from 1996 at The University of Chicago Press (Chicago 1996)).

Flusser, Vilém (1996 [1985]). *Ins Universum der technischen Bilder* (Ed. by Andreas Müller-Pohle, 1. Auflage 1985). Göttingen: European Photography.

Flusser, Vilém (2011 [1985]). *Into the Universe of Technical Images*, (Electronic Mediations Volume 32, introduction by Mark Poster, translated by Nancy Ann Roth). Minnesota / London: University of Minnesota Press.

Ketelaar, Eric (1997). The Difference Best Postponed? Cultures and Comparative Archival Science, *Archivaria. 44*, 142-148. (Reprinted In P.J. Horsman - F.C.J. Ketelaar - T.H.P.M. Thomassen (ed.) (1999), *Naar een nieuw paradigma in de archivistiek. Jaarboek 1999* (pp. 21-27), Den Haag: Stichting Archiefpublicaties).

Ketelaar, Eric (2001). Tacit Narratives: The Meanings of Archives. *Archival Science, 1*, 131-141.

Mcluhan, Marshall (1994 [1964]). *Understanding media: the extensions of man*. Cambridge, MA: The MIT Press.

Zizek, Slavoj (1996). Cyberspace, or the Virtuality of the Real. *Journal of the Centre for Freudian Analysis and Research (JCFAR), 7*. Retrieved from http://jcfar.org.uk/wp-content/uploads/2016/03/Cyberspace-and-the-Virtuality-of-the-Real-Slavoj-Zizek.pdf

W O L F G A N G   E R N S T

Technologies of Tradition: Between Symbolic and Material (Micro-)transmission[1]

Introduction: Cultural analysis in the media-archaeological way

The analysis of the mechanisms of cultural tradition is a subject which traditionally requires training in historical studies. The following argument claims that there is an alternative way of approaching that subject: media archaeology as a research method. Inspired by Michel Foucault's *Archéologie de Savoir* (1969) but being more precise in extending this approach to material and technological culture, media archaeology does not primarily locate cultural phenomena by discourse analysis, but seeks to unreveal the Foucauldean *archive* of cultural knowledge in the grounding (German *Erdung*, a technical term in electro-engineering) in both material and logical artefacts (see Parikka, 2012). Media archaeology stratigraphically discovers a layer in cultural sedimentation which is neither purely human nor purely technological, but literally inbetween (Latin *medium*, Greek *metaxy*): symbolic operations which turn the human into a machine as well as they can be performed by machines.

In comparison with classical archaeology, media archaeology shares the interest in material culture. There is a specific affinity between media archaeology and the analysis of cultural engineering. What differentiates technological objects from archaeologically excavated cultural artefacts from past civilizations is their essence as conceptual circuit diagrams and source codes. Technological objects come into being (technically as well as logically) when being biased with electric current – contrary to a simple museal assembly. This escalation is articulated in the different emphasis between a historiographical "history of things" and a media-epistemological analysis of micro-physical materialities (Lubar, 1993; Galison, 1997).

As an emergent, cross-disciplinary field of epistemological inquiry and lab-based experimentation debate, media archaeology is still open for ongoing redefinition. Some of its practitioners agree in an interest to revisit and reconstruct media from past cultures with particular attention to obsolete and outmoded media technologies and cultural engineering practices, for the sake of challenging some of

---

[1] This text originates from the following keynote speech, but was substantially rewritten: "(Micro-)Transmissions. Technologies of Tradition", Keynote speech to interdisciplinary Ph.D. student seminar *Materiality and Historicism* of the Norwegian research school "Text, Image, Sound, Space", Centre Franco-Norvégien en Sciences Sociales et Humaines (FMSH), Paris, 25-27 January, 2012.

the dominant metahistories of technological progress that still shape our view of the past in the present. Other media archaeologists even question the plausibility of narrative itself as plausible assumptions for describing the being of technologies in time.²

In a rather rigid version of what is media archaeology, its theoretical stakes, and its recommended *modus operandi*, this text argues in favour of a radically material and mathematical approach to the study of cultural change, memory, and knowledge tradition, and even the very category of time itself. Such an approach displaces the human subject as the central figure of cultural history, and explores instead the role of non-human, non-discursive infrastructures and technical practices that shape how the past is made present, how time is recorded, and how the documents and narratives depend on technical methods of preservation and organisation.

Embracing a variety of technological *a priori*, the reader is therefore invited into a world that is understood to exist almost independently of the literary frameworks of meaning through which the past and present are usually bridged: a universe of oscillatory vibrations and rhythmic pulsations produced by the machines and the techniques that human cultures have relied upon to divide, measure, and store intervals of time, to harmonise sounds, to map space. Even the apparent anthropologically fundamental (or even metaphysical) separation between presence and absence has been redefined by technological means of differentiating between "one" and "zero", a distinction that enabled the binary code of modern computers, but that can be traced back through longstanding historical precedents, such as the mechanical hindrance in clocks since late medieval Europe. Media archaeology pays attention to such pre- or nondiscursive elements, which are embedded in particular techniques and instrumental means of organising the basic building blocks of knowledge, such as temporality, presence, arithmetic, geometrical modelling, harmonics, and so on. Media studies must be prepared to break with narrative historiography and confront machines on their own terms, to engage with them truly media-archaeographically. This is not simply an abstract, theoretical exercise; media theories only can be valued only when being tested against hard(ware) evidence. Media scholarship – to the extent that it depends on the existence of archives of documents (codes) and material culture (hardware) from the past – needs to take this radical questioning of its own technological bases serious.

In principle, media archaeology is not just insisting on the materiality of media in culture, but also revealing the power which drives them nowadays – which is algorithms, mathematical tools, in the precise sense of the ancient Greek "*arché*" which does not mean origin in the historistic sense but traces preconditions of cultural practice like the mathematical square root (media archaeology is "radical" in that sense). This prevents media archaeology from being simply nostalgic in terms of hardware and adds a sharper edge: the mathematical analysis.

---

² My paraphrasing of the media-archaeological method explicitly refers to the opening address given by Jeremy Stolow to my lecture "Temporalizing the Present and Archiving Presence. The impact of Time-Critical Media Technologies" at Concordia University, Montreal, September 27, 2014

This confrontation of the physical, time-based signal with logical, negentropic information poses a challenge for historicism in the age of digital media, resulting in a double bind of historical and ahistorical tempor(e)alities. Martin Heidegger frequently differentiated between history as lore (German *Historie, Kunde*) in the vulgar sense and historicity (German *Geschichtlichkeit*) as (trans-)mission of being itself: a trajectory into the future rather than a simple reference to the archive of the past.[3]

A sense of discontinuities, ruptures, thresholds, limits and series – in Foucault's sense – separates media archaeology from the traditional model of writing media history. For a mind trained in occidental culture, it is unusual to rethink multiple temporalities in ways which do not reduce any event in time to history. Trying to resist a reductive, linear narrative of technologies of tradition which would start in the past and result in the present condition ("from" materiality "to" the virtual), this argument intends to analyse both the material (entropic) forms of cultural tradition and the immaterial, almost time-invariant codes of transmission (the physical and the symbolical mode, material embodiment and logical implementations), elaborating on the shift from archaeological materialities as cultural premise to techno-mathematics as the new form of enculturation.

Re-thinking cultural historicity: Transmission across space / tradition over time

There is a specific historicity which is embodied or implemented in the materialities of communication (Gumbrecht, 1988) – split into technologies of transmission (space-centred) and tradition (time-centred). There are different tempor(e)alities at work. When two separate partners (in fact: e-mail programs) communicate *online*, information passes through space, while "when a program stores data in a file (or a person saves a notebook in a desk drawer), the intention is to pass information through *time* <...>." (Gelernter, 1997, p. 100).

The Canadian economist Harold A. Innis, one of the main inspirators for Marshall McLuhan's media theory, in his monography *Empire and Communications* (1950) differentiates between space- and time-accentuating forms of transmission, depending on the materialities (if not "media") of communication. As functionally embodied in such media, empires are either spatially or temporally "biased".[4] According to Innis, inherent to the material properties of cultural technologies are spatial or temporal determinations which are interrelated to modes of executing power.

Objects distributed over space allow for immediate communication, objects distributed over time favour persistence. But both modes are not categorically distinct but extreme formulations of one and the same dynamical eventuality. With advanced physics and neurology we get used to the idea that both Kantian *a priori* of

---

[3] On temporality ("Zeitlichkeit"), see Heidegger, 1998, p. 136

[4] "Bias" originally is a technical term in electronic engineering describing the necessary current polarisation to operate a vacuum tube – a literally pre-conditioning, a ground tension, an electric *a priori*. For magnetic recording, the "bias" names the pre-magnetisation of the tape by high frequency signals to ameliorate the signal-to-noise ratio (dynamics).

perception (time and space) are not that distinct at all, but rather relativistically interlaced (like the dynamics of the electromagnetic "field").

The Internet today corresponds to the principle analysed by Innis – the "Carthage" option which means hyper-linked exchange in ports across the sea, as opposed to the land-based Roman territory (empire). Here the act of transmission is the message (*tele*communication / *telos*) – exhausting itself in operation, in contrast to long-time teleological monumentality as embodied by the Egyptian pyramids.

In the age of transnationally networked economies and cultural exchange by technological media, the term "tradition" is subject to a dramatic change of meaning. Emphatic macro temporal ("historical") tradition is being replaced by time-based and time-basing micro-mechanisms of transmission. While tradition has been associated with long-time memories across deep historical time so far, this emphatic horizon now seems to shrink to a mere extension of the present (as its re- and protentive short-time "working memory") – a dramatic shifting of the temporal prefix.

The expression "cultural transfer" already implies a causal relationship, a traceable chain of the channels whereby cultural goods as materialities or cultural knowledge as information are negotiated between cultures. By using the term "transmission" instead, a technological co-significance comes into play which has arisen with the epistemology of electric and electronic media since nineteenth century telegraphy: the idea of "wireless", that is: immaterial transmission of signals. Whereas this corresponds with the sender-receiver model in communication engineering, an alternative option is a theory of co-original emergence of structurally similar technologies without direct exchange of knowledge, according to the principle of the communicating tubes (and resonance). The development of printing with moveable type in ancient Korea already took place before (and independent of) the Gutenberg press, and principles of the wheel-driven clock were developed in China before it emerged in late-medieval European monasteries – without mutual knowledge transfer.

Archival materialism versus streaming electrons

High-technological media culture is characterised by the dichotomy between materiality based endurance and electronic immediacy. Emphatic transmission over time ("tradition") is based on the materialities of the archive. On the other hand, immaterial electronic "live" media (transmission across space), by definition (and essentially) are rather memoryless. Different from the legal record which is tentatively meant to last for eternity, the quick notice on a random access data carrier (starting with the ancient wax tablet) already fulfils its function as intermediary storage. Thus, a new kind of evidence emerges: fugitive moments of the past which were never meant to enter the discourse of history.

An example is the "direct recording": the immediate engraving of phono discs in the 1930s and 1940s. The material recordings were not used for long-term storage but for short-time delayed replay in broadcasting across temporal zones or on different program places, since magnetic tapes were not yet available. Their function was

fulfilled (actually exhausted) in the act of intermediary storage – just like the intermediary film with washable emulsions on celluloid in early television transmission of outdoor events by daylight; emphatic tradition here is replaced by delayed transfer. It would be a misunderstanding to categorise them under material storage media; they were rather intermedia in the temporal sense (the time channel). The more informative they are when, by accident, they were not effaced but frozen over time, to be discovered in the present. Their information value is reciprocal to their (non-)intention as "historical" records.

We can observe a transformation of an epistemological dimension: the transformation of the classical, datacarrier based, material storage "archive" into a literally e-motional archive (that is: in electronic motion), in electromagnetic ephemerality and latency. The gain of flexibility and computability is paid with a loss of durability.

When recently the Cologne Municipal Archive materially collapsed, it became apparent that most records, though being dirty and mutilated, materially survived this catastrophe, astonishingly resistible against the pressure of stones. In a similar way the first generation ("analogue") audio-visual storage media turned out to be surprisingly resistant against temporal entropy (like the Edison-cylinder and gramophone records, as well as daguerreotypes, photographic negatives and film on celluloid). More delicate is the destiny of cultural memory based on electromagnetic storage; digital media, finally, tend to divest themselves completely from their material embedding – losing the "touch ground" by becomingtechnically "virtual".

Traditional physical storage media have been materially inscribed (*graphein* in its old Greek sense): "There must be a writing means by which the information to be stored is introduced into the device" (Sage, 1953, p. 141-149 (141).; against this invasive inscription, storage devices such as magnetic tape for audio and video only reveal their memory content in the dynamics of the electro-magnetic field. Electro technical storage media take place in a sphere which is different from the scriptural regime of the classical archive – until, on the level of alphanumeric codes, alphanumeric writing unexpectedly returned within techno-mathematical machines. This return is a temporal figure which cannot be reduced to the linearity of media history; as a Möbius loop, we are confronted rather with a media-archaeological (re)configuration of times present and times past: a contemporalisation.[5]

There are two complementary approaches to the conservation of analogue memory carriers. The one cares for preserving the physical, especially chemical and electromagnetic properties of the concrete media body – since all media technologies are hardware in the first place. The other, sometimes opposing approach is to preserve media-based memory as information, up to the extreme point of view that the material body might be abolished after its essential transformation into its pure binary information units.

---

[5] To experience the past *both* in its characteristic difference and at the same time as present in actual co-presence is a double-bind proposed by T. S. Eliot, 1949, p. 64

In his discourse-generating publication *Cybernetics* (1948), Norbert Wiener provocatively declared that information is neither matter nor energy. Translated into the question of material tradition in culture, this means that archives no longer collect the carriers as storage media of inscriptions but simply the information itself. To which degree does the authority of an archival record still depend on its material physical embodiment? Is it no longer important by which carrier one generation passes on its information to the next?

For the oldest analogue signal-based medium in the technical sense, photography, in 1859 Oliver Wendell Holmes pointed to the fact that this symbolic trade of media and material was introduced by photography: "From now on, form is separated from material. In fact, the material in visible objects is no longer of great use, except when being used as a model from which the form is constituted. Give us a couple of negatives of an object worth seeing … that's all we need. Then tear the object down or set it on fire if you will … the result of this development will be such a massive collection of forms that it will have to be arranged into categories and placed in great libraries." (Kemp, 1980, p. 121)

Is there (in the name of information) a transcendence of media-material in photography, undoing historicity? Transcendence of materiality is a cultural effort well known from religious practice (e. g. the art of orthodox icon painting in Russia). In medieval Europe, light in cathedrals was meant to transcend the material boundaries of architecture. With photographic emanations, light itself becomes a "historiographical" index (or even media-phenomenologically transcends history by its affect of immediacy on the human temporal sense: preserving the past as present). But still this is not immaterial but bound to a chemical storage medium. *Temporal* transcendence of materiality is a faculty of operative media technologies.

Once the signals which are chemically fixed (photography), mechanically engraved (phonograph) or magnetically embedded (magnetophon, videotape) on a material carrier have been transformed into digital, immaterial information, they can be (virtually lossless) "migrated" from one storage computing system to another. Permanence and archival endurance is not achieved in the traditional way anymore (which has been monumental fixation, *stasis* so far), but by dynamic refreshing – the "enduring ephemeral" (Chun, 2001, p. 184-203).

Material entropy

Let us apply the notion of "media materialism" not just in a neo-Marxist sense but in its pure physical meaning. "Tradition" on the very material level is the ability, f. e. of chemical polymers, to retain the "memory" of their formal juncture. But magnetic tape in audiovisual archives is a fragile medium; the "Vinegar Syndrome" (the chemical disintegration of the carrier material) cannot be counter-chemically stopped, just slowed down. On the other hand, the structure of the containing archive is negentropic: a symbolic order. Both converge in so-called "digitisation", where the media carrier in its physical entropy no longer counts, but the content (signal) as information, stored in the symbolic (binary) code.

The microchemical analysis of such "endangered" material, by Fourier Transform Spectrography, is a very close, truly media-archaeological reading of such archival media materialities. Such microtechnologies are at work in what discourse emphatically calls "cultural heritage" embodied as material tradition. When the notion of "the material" becomes dynamic itself – as identified by Henri Bergson[6] –, historicism is cut short to the atomic level. Contemporary with Albert Einstein's development of relativity theory, around 1900 Henri Bergson elaborated his philosophy upon the physical insight into the essentially oscillating nature of "material" elements (electrons, atoms); as such, materiality itself is in micro motion already. This becomes evident in electronic music. What used to be an essential element in music composition, the tone, turned out to be a microtemporal event in itself – the tone as a frequency event.[7] With electronic signal processing, the traditionally separated categories of durable materiality *versus* immaterial time-based performance collapses.

Negentropic information of matter

Material media have their individual entropy, characteristic probabilities of physical endurance – *Eigenzeit*. The physical media differ from the software-based media by embodying fundamentally different temporal regimes.

Physical (entropic) decay of cultural materialities from the past may be diagrammatically registered and presented like a signal, that is: as a continuous function of the time axis (e. g. a building, slowly degrading into a ruin or finally a heap of bricks). The alternative temporality is discrete, a sudden change of state which then remains invariant over a long period of time (interval) – close to PCM in digital data transmission (based on impulses rather than on continuous signals). This results in a dramatic increase of data compression, since only the changes (between two states) have to be coded. Thus, the physical time axis is being replaced by coded time. Transmission itself (in its traditional sense) runs out, becoming a mere function of mathematised (rather than materially transmissional) signal processing, achieving real-time effects by compression algorithms (Siegert, 2003, p. 285). Compressed time takes place when digital imagery is being transmissed, notably in compression algorithms as defined by The Moving Picture Expert Group (MPEG standards for streaming digital video), whether recorded – from the archive – or in "live" transmission (Richardson, 2010).

For the longest time in human history, the notion of media-based transmission was bound to the material carrier; tradition thus appeared as a kind of generalised mailing service not for the conquering of space but of temporal distance. This has changed in the époque of symbolic information processing. Leon Battista Alberti in the Renaissance, in his treatise *De statua*, already proposed a procedure for the lossless transmission of three-dimensional objects by digitalisation. Once a human body or an artefactual device has been subdivided into a grid of discrete points, the position of each one of these points can be precisely indicated by a system of spatial coordinates. Such an alphanumerical list enables the original to be copied and equiprimordially be regenerated *ad infinitum*, be it in distant places or in future

---

[6] Bergson's concept of matter has been re-interpreted for the electronic image by Lazzarato, 2002

[7] See Stockhausen, 1963: "Die Trennung 'akustischer Vorordnungen' *im* Material und 'musikalischer Ordnungen' *mit* diesem Material müßte dann aufgehoben werden." (p. 214)

times (Carpo, 1998). Such a procedure avoids the distortions which happen in the act of manual copying, resulting in a model for lossless, negentropic tradition by means of coding the image, that is: its informatisation. From that practice results an almost ahistorical form of tradition, which manifests itself not only in the current age of born-digital records and is known from the archival rescuing projects of digitising endangered historic manuscripts, but turns out to be the essence of communication engineering.

Informational historicism: Data-archaeological reconstruction

The Frauenkirche cathedral in Dresden from the Baroque times endured more or less unchanged into the twentieth century, as documented in countless paintings, engravings and photographs). After the great fire in Dresden in February 1945, resulting from the serious bombarding of the city by British airplanes in World War II, the building (with a slight delay of two days after the actual bombardment fire) collapsed (having been burned from the inside), from then on remains more or less unchanged as an almost defigurated heap of stones. Nowadays, with the aid of computing, the cathedral has been reconstructed out of the (available) original stones, reversing the collapse which happened after the Dresden bombardment – a negentropic violation of the second law of thermodynamics defining entropy which provides the scientific explanation for the notion of a time arrow. *Informational* aesthetics informs matter itself.

The historicity of "materiality" in the age of information technology

No cultural artefact can be reduced to sheer materiality but rather needs to be differentiated into *physis* and *logos* – the very meaning and double-bind of "techno/logy", known since the Aristotelian distinction between *logos* and *physis*.

Logos is expressed in symbols, which as cultural engineering is linked to alphabetic writing. In contrast to the orally transmitted tradition, "writing <…> yields a 'transpersonal memory'. Historically it has facilitated abstract thought, giving rise to science and mathematics, and <…> freed thought from the subjective realm" (Heyer, 1995, p. xvii) – a structural, symbolic regime, close (but not yet completely, that is: automatically emancipated) to a *non-human agency* (as described by Bruno Latour). In the époque when Charles Babbage extended his arithmetical *Difference Engine* to a storage-programmable, thus algorithmic *Analytical Engine*, it was G. W. F. Hegel who opposed the idea that the act and procedures of thought might be (as expressed later in Boole's *Laws of Thought*) performed by a logical machine rather than by «working through» in philosophical terms. Likewise, Hegel differentiated memory (*Gedächtnis*) as cold mechanical storage (such as library systems) from the interiorisation as remembering (*Erinnerung*) which corresponds with historical imagination.

Within the digital computer, calculation and intelligence converge: the mechanisation of spirit and the spiritualisation of matter coincide (Kittler, 1989, p. 30f). When the computer is seen disintegrated in an electronic waste deposit, it is indistinguishable from other fragments of ("analogue") electronics. Its specificity is not in its hardware materiality, but in its algorithms (which might be performed in

organic matters as well, such as the DNA computer). If future archaeologists will discover strange artefacts within the ruins of our present, these will obviously be interpreted as primarily electronic devices, and they probably miss the essence of computers as symbolic machines.

Software: Towards a new meta-physics of preservation?

For a long time, memory correlated with monumentality. The endurance of a traditional "archival" record (document) used to depend on its material storage medium, with the option of "migrating" symbols *as information* from one carrier medium to another. The Roman poet Horatius derived his post-mortal optimism from this:

> "Exegi monumentum aere perennius
> Regalique sita pyramidum altius,
> Quod non imber edax, non aquilo impotens
> Possit diruere aut innumerabilis
> Annorum series et fuga temporum."*
> (Horatius)

Medieval monks copying an ancient manuscript for tradition either involuntarily (carelessness) or involuntarily (the desire to correct the source) made mistakes (Nita, 1998, p. 402). Still, the symbolical code can be transmitted with a high degree of fidelity in copying, regardless the material support. Thus, the symbolic code (like the genetic code), especially the phonetic alphabet, is mostly invariant towards historical, i. e. entropic time. Digital data is just a special case of such alphabets.

Documentary science has developed the notion of "logical preservation" (Marker, 1998, p. 296). But any information must take place in or on a material support which introduces another, different tempor(e)ality. Does the concept of "information" (which is measured by the binary digit) dispense with the material link? To what extend is software independent of the carrier used for transport? (Swade, 1998, p. 195). In order to be executable, any algorithm has to take place in matter – even if it is just paper. The metonomy which takes the Floppy Disc as a material support for the software itself is a hint to the material link.

If past information is not just functionally emulated but actually simulated, its temporal (entropic) behaviour must be archived as well – like the scratch, the noise of an ancient Edison phonographic cylinder when being digitised. One method is known from computing as physical modelling.

A piece of software is a set of formal instructions, or, algorithms. "It doesn't matter at all which particular sign system is used as long as it is a code, whether digital zeros and ones, the Latin alphabet, Morse code or, like in a processor chip, an exactly defined set of registers controlling discrete currents of electricity." (Cramer, 2002). Suddenly, hardware itself can turn out to act as software.

---

* "I have created a monument more lasting than bronze, / And higher than the royal site of the pyramids, / Which neither harsh rains nor the wild North wind / Can erode, nor the countless succession of years / And the flight of the seasons" http://home.golden.net/~eloker/horace.htm

Is algorithmic knowledge better preserved when transmitted by an exclusive expert group (software "priests", according to the monopolistic Egyptian tradition of knowledge transfer) than by printed publications?

The Department of *Computing and Control* of the National Museum of Science and Industry in London, faced the challenge of the preservation of software as museum object. Software is a new kind of cultural artefact: it is not a material object anymore, but rather an executable file which unfolds only when being processed (a truly processual time object). A computer as hardware can be traditionally displayed as an immobile object, but its time-critical and bit-critical processes are never in *stasis*, just like frequency-based acoustics (sonic evidence in museums) need performance in time to take place – different from visual evidence which persists in space.

Software belongs to the class of generic objects media. "One bit wrong and the system crashes"; therefore "in archaeological terms the operational continuity of contemporary culture cannot be assured" (Swade, 1992, p. 208f) as soon as the material embodiment in which such a software must take place in order to actually run is not available anymore. The solution to this material dilemma lies in transforming the material aspect of computer culture itself into software, that is: emulating past hardware digitally. Suddenly cultural tradition turns out to be an operation of dematerialisation (German *Verundinglichung*), "logical replication as distinct from physical replication" (Swade, 1992, p. 209). In fact, operational media actually transcend material "things".

The evolutionary biologist Richard Dawkins transformed the notion of the "gene" into the "meme" as an agency of cultural transmission, turning humans themselves into channels of knowledge transfer. (Dawkins, 1989). Information thereby replicates in the very act of communication over time and space – up to the World Wide Web and the "viral" media sphere of today.

Archaeology versus history

It is not by coincidence that one of the first sciences in the humanities department which applied computing has been archaeology. This is not by chance but reveals as structural affinity. Archaeology, especially (appropriately so-called) prehistorical archaeology, deals with pure material data, no narratives (textual tradition) like the classics in Greek and Roman philology. In many ways, archaeology is close to mathematics. Epistemologically, this becomes clear with Michel Foucault's discourse analysis in terms of propositional logic (Kusch, 1989).

As opposed to figurative, narrative (hi)storytelling, archaeological processing of past data concentrates on what used to be familiar as "antiquarian" modes of representation under the auspices of digital computing (Ginouvès, Sorbets, 1978). This leads to *dia*grams (with the prefix indicating a temporal vector) instead of historiography. The implosion of the narrative frame has consequences on the form of representation of the past. Instead of being governed by the apparently seamless and unbroken literary text, figurated and effected by rhetorical moves and dramatic emplotment, modular writing is governed by the non-discursive logistics of vector

fields, graphically expressed by means of marks of the directing codes <*Steuerzeichen*>: networking rather than narrating the evidence. Hermeneutics itself becomes algorithmic:

"Archaeological *data* consists of recorded observations. These might be measurements of the size of a hand axe, the stratigraphical relationship between two layers or the geographical location of a site <»Ephesos»>. Whilst archaeological data is frequently numeric, it can equally well be non-numeric, such as the name of the material or colour of an object. It also comprises visual data, such as photographs, plans or maps. Data *processing* is the name given to the manipulation of data to produce a more useful form, which we shall call *information*. <…> The sequence of operations required to perform a specific task is known as an *algorithm*." (Richards, Ryan, 1985, p. 1f)

Let us distinguish diagrammatic *archaeography* from more interpretative archaeology in a narrower sense (Moberg, 1971, p. 533). The philological practice of constructing genealogical filiations of manuscript tradition in the form of *stemmata* applies a diagrammatic method.

On the other side, there is data processing as archaeology. Media archaeology is not just a way of remembering "dead media", but rather a mathematical aesthetics; modelling, statistics and especially cluster analysis (e. g. for the distribution of objects in a grave field) is one the fields where archaeology made use of data processing with electronic computers. All of a sudden, the memory of material culture becomes related to mathematics instead of *belle lettres*. Mathematical methods (like stochastics in "cluster analysis" of graveyards, f. e.) are being applied in archaeology.

"Writing vs. Time": Lossless tradition, message or noise?

In every act of cultural transmission, there is a symbolical (code) level on the one hand which is time-invariant, and an entropic, temporally decaying ("historical") physical real(ity) on the other. Let us take as an example for symbolical tradition the transmission of Euclid's *Elementa* from Greek antiquity to the European Renaissance *via* Arabic translation (intermediation). Here, the name (the medium) is the message: *Elementa* is the name for letters (the ancient Greek alphabet) and numbers, which serve as the concrete symbolic medium of transmission. The subject of this work, mathematical geometry, itself claims metahistorical truth (the Platonic anamnetic knowledge), while the physical embodiment of this symbolic knowledge, f. e. ancient book rolls, are subject to decay.

Techno-implicit knowledge[8] traverses cultural history according to temporal laws of its own – it is self-repetitive, close the model of "memetics" (a kind of cultural memory gene, as defined by the evolutionary biologist Richard Dawson).

Let us look closely at a painting which sums up these conflicting energies of tradition: Anton Raphael Mengs' painting *Allegory of History* (1772/73) on the ceiling of the *Stanza dei Papiri*. It links the Vatican Library with the Vatican

---

[8] Different from Michael Polanyi's notion of "implicit knowledge": see Polanyi, 1958.

Museums topologically, and thematically represents the dichotomy between material and symbolical objects and records of cultural transmission: physical entropy *versus* symbolical (ahistorical) invariance.

Museums, libraries and archives – all three memory agencies which act in the Vatican context of Mengs' allegory – are agencies of cultural transmission across time. The dramatic setting of Mengs' allegory is about conflicting tempor(e)alities which are at work with cultural tradition: Chronos (physical, material entropy) *versus* Clio (symbolic coding). In this allegory, storage and transfer media are not just rhetorical metaphors for cultural tradition. In fact, technologies of tradition are literally *metaphorical* (Röttgen, 1980, p. 121).

Mengs' *Allegory of History* features a *genius* who is transferring papyrus rolls to the personification of history (Clio), in fact performing the *archival* act: which is the rescuing of physically endangered records from the past by transcription into symbolic historiography. The material record from the past (subject to entropy) is thus translated into (negentropic) information. Such an act of transformation is well known from the current massive digitisation of, f. e., early sound recordings in the phonogram archives of Vienna and Berlin – a chrono-economical exchange between the real and the symbolic, between aging and permanence.

Mengs' *Allegory* demonstrates the authority claim of the Roman church which is based on long-time tradition (*monumentum*). The status of the museum objects depicted on the painting is both material and semiophoric, depending on their internal or external relation to the subject – the allegory of the *Museum Clementinum*. Two regimes conflict here: registering and description, *versus* historiographical narrative. On the borderline between history and archaeology, it is not clear what Clio performs in the museum: does she write or register? Her attention is diverted by double-faced Janus who points at the realm of the aesthetic (represented by the Cleopatra/Ariadne in the museum), whereas in fact what is brought to her is data. Instead of historiography, her book might be an inventory.

The allegorical figure of Chronos embodies the physical reality of time which is entropic decay. Asymmetrically, historiography is embodied by the female allegory of Clio who records chronological events in the rather time-invariant symbols of writing. But the chronological order (counting historical time) is a historicist distortion of temporality itself: "Use of centuries – fingers and toes – distortion of history", Harold A. Innis 1947/48 wrote in his Idea File.

Subject of the Stanza dei Papriri – both in its archaeological content and its painted allegory – are the cultural technologies of transmission in time in their various forms. When Mengs painted an ancient inscription in this scene, it had just been interpreted as the donation mark of the family tomb of T. Claudius Primigenius <comp. "Genius", der die Papyrusrollen heranträgt, und der in der Philadelphia-Version auf die Inschrift schaut!> who had been archivist of the imperial domains in ancient Rome. Thus, the epigraph doubly offered itself (by its donator and by its function) as a welcome supplement to the overall theme of the fresco: The archivist as gatekeeper of historical memory here cares for his own remembrance. In a

previous design for that fresco (preserved in the Philadelphia Museum of Fine Arts) Mengs lets Saturn (Chronos) look directly at the inscription which in its materiality reveals apparent traces of decay and age (Röttgen, 1980, fig 10).

Represented here is a literally archaeological moment, the excavation of an ancient inscription – not simply as an allegory of vanity <comp. Winckelmann's critique of baroque Ruinenmelancholie>, but as a symbol of rescuing the heritage of pagan culture by means of the storage place museum.

There is another eighteenth century allegory of the mechanisms of transmission, the frontispiece of Lafitau's publication *Moeurs des sauvages Ameriquains* (1724). This image confronts archaeologically silent, but enduring material artefacts with the discursive, but transient articulations of historiography. The viewer is confronted with the encounter of writing and time in a collection space littered with material traces coming from both Classical Antiquity and the New World:

"One holds the pen, the other the scythe, […] which approach each other without ever touching, asymptotically. History deals with relics which can be seen, and seeks to supply explanations; ancient *things* which have become mute through the degradation owing to time may to some extent become clearer if we invoke *customs* observed among contemporary savages. This operation needs […] a technique, which is that of comparison […] and an author, an historian." (Lavers, 1985, p. 330f)

Archaeology deals with gaps and therefore faces traumatic absences; historical discourse is made to fill this up to generate some kind of symbolic order on the material ruins of tradition.

Michel de Certeau enhanced his interpretation by drawing the configuration of Chronos and Clio abstracted to a diagram where the supposed prologued lines of the curved scythe and the linear pen become vectors. Directly deciphered in terms of mathematics, the pen-line (as x-axis, the abscise) becomes the asymptote of the scythe as hyperbole (on the y-axis). There is no point where the function touches or traverses the x axis itself: no convergence between material ("historic") and symbolic ("historiographical") phenomena of time.

In Lafitau's front cover illustration, the allegorical figure of Chronos is endowed with a weapon (the scythe) indicating devastation with time – in fact "noise" which happens in the temporal channel of transmission (to rephrase it in terms known from transmission engineering); such material loss of information is compensated by the female allegory of Clio "writing" history: copying of symbolic letters is an almost lossless technology of tradition. A different loss though takes place at the moment when real matter or energy is symbolically filtered, that is: compressed.

Tradition here means the separation of signal from noise by means of symbolic transcription. When a message has been received which has somehow become scrambled with another, unwanted message (usually called noise), the challenge lies in "unscrambling these and restoring the original message with as little alteration as possible, *except perhaps for a lag in time*" (Wiener, 1948/50, p. 205, italics W.E.) –

which is the problem of filtering. But an ancient clay tablet from Babylon which copies an even older astronomical text already notates the material corruption of textual evidence in the original – with the word *hepi* (meaning "brocken") (Hunger, 1976, p. 47ff on clay tablet SpTU I 94).

In fact, cultural tradition can be rephrased in terms of engineering. Claude Shannon's techno-mathematical theory of communication (1948/49) which concentrates on media channels can be extended to the mechanism of emphatic cultural tradition as such.

"We may assume the received signal E to be a function of the transmitted signal S and a second variable, the noise N. <…> The noise is a chance variable just as the message <…>. In general it may be represented by a suitable stochastic process." (Shannon, 1963, p. 65)

What historians hermeneutically narrate as probability of tradition (Esch, 1985), can be mathematically expressed exactly. The time axis is the diachronic dimension for the transmission of encoded signals; such a technically informed semiology does not distinguish meaningful chance from random events (accidents, noise). But for historians, only the message, and not the noise created in the transmission channel, is taken into historiographical signification (and thus limited to what can be expressed by the alphabet). Looked at in a media-archaeological way, the apparent noise might be the arbitrary acts of encrypting another message.

Towards a mathematical theory of memory communication

Analog and digital communication, based on continuous signals or discreet symbols (like telephone talks and archival readings) can be formally expressed as transmission, be it sequences of dots and dashes (the Morse code) or wave patterns (telephone, phonograph, analogue electronic television). The notion of cultural transmission already implies an intentional act, an addressing of posterity. "For communication as we discuss it here, the addressee is almost always a human recipient in a multiplicity of channels – whether directly, as in film or on the telephone, or indirectly as via a postcard or a secret code, <…> across space (as in radio or television) or across time (as in books or audio recordings) <…>." (Krapp, 2011, p. xii) When the historian (researching in the archive which is one such secret codes) places himself as the receiver, this is an act of supposition. The term «sending» here can be understood not as destiny in a metaphysical way but as a concrete act of mailing, corresponding as an act of transmission engineering with what Walter Benjamin has called the "historical index". Images from the past may be indexed with a kind of implicit time code:

"The past 'carries with it' a temporal index: the date of its emergence and of its expiration. [...] The address of the past in all its power *will have been* if it is read by the present that it enables; if it is not, it disappears without a trace. [...] Benjamin always thought the address of truth in historical (or at least temporal) terms; translatability, after all, comes about only in time and for a time, and translation is not a mere transcription." (Fynsk, 1978, p. 577 f)

Intended for tradition, records from the past are endowed with addresses (to posterity) which implies a (virtual) notion of the organisational archive already, as opposed to random transmission of past remnants which is noise rather than message. Noise belongs to the kind of signal that the sender does not want to transmit – a situation the archaeologist is most probably confronted with. Transmission by symbols here differs fundamentally from the endurance of material artefacts.

Mathematised electronics (treating signals in terms of information) is able to detect, filter and regenerate digital pulses in spite of distortion or noise in the channel – e. g. by application of the *repeater-regenerator*, as a medium of almost time-invariant transmission. The binary nature of the digital undercuts the well-known parameters of historical tradition which are subject to informational loss over time; a binary signal can be obtained even under conditions of heavy interference as long as it is possible to recognise the sheer presence of each pulse almost independent of the length of the channel.

What today is being decoded as "message" from the past can be mathematically formulated as a function of its signal-to-noise ratio. Let us add the temporal dimension (the time axis t) which serves as the channel of communication between past and present. From an engineering point of view, communicative transmission and cultural tradition interfere, as expressed in 1888 by the inventor of magnetic sound recording, Oberlin Smith: "Imagine that speech could be transmitted over a telephone line at a very slow, rate of travel, so that at a particular point in time the entire message would be somewhere in the wire between speaker and listener" (Engel, 1986, p. 171) – which is in a literally *medial* state of existence. Shannon defines the channel of transmission as "the medium", which corresponds with Aristotle's ancient definition of "the inbetween" of communication (*to metaxy*).

Understanding the past by resonance instead of historicity

Among the channels of media-internal communication within early electronic digital computers, there have been short-time dynamic storage devices such as ultra-sonic mercury delay lines. Media-archaeological imagination feels tempted to correlate this form of intermediary storage with the temporality of cultural transmission. This results in the assumption that it might be possible to listen to the sound of tradition as soon as human perception is tuned to resonate with such vibrating waves and impulse trains, at the borderline between the physical materiality (endurance) of the past and the tempor(e)alities of historicism.

Is there something like immaterial communication across time? The cultural historian J. Bachofen once remarked that when we imagine ancient Rome, a momentary flash like an electric spark immediately springs from Roman antiquity to the present, undertunneling the historical distance in between – a kind of radio communication across time:

There are two roads to every kind of knowledge, the longer, slower, more laborious one of intellectual combination, and the shorter one, the one we cover with the

energy and speed of electricity – the road of the imagination when it is touched by the sight and the immediate contact of ancient remains and grasps the truth in a flash, without any intermediate steps. (Gossman, 1983, p. 49).

The antiquarian approach to the past is by taking the archaeological materiality remaining from the past at face value: "There is something about the walls of Rome that moves the inmost depths of man. When a metal plate is struck, the iron resounds and the echoing is stopped only by laying one's finger on it. In the same way, Rome moves the spirit that is in communication with antiquity … all that was slumbering within him." (Bachofen, as quoted in Gossman, 1983, p. 46f)

Knowledge transfer is dependent on the receiver to resonate with its carrier oscillations; by converting analogue waves into symbolic frequencies, they can be accessed by mathematical reasoning (Erlmann, 2010). Aby Warburg in his concept of persisting visual gestures in occidental cultural history refers to sub-cultural *mimesis*, somewhat replacing the notion of diachronic tradition by the notion of almost immediate transmission between sender and receiver when tuned to each other (Warburg, 2010, p. 640) – a model rather derived from the engineering of wireless communication than from historical hermeneutics.

Material philology: The Case of Lapis Satricanus

What happens if cultural heritage is subjected to the microscopic gaze? The title of a German publication from 1973 on the scientific (rather than hermeneutic) analysis of art historical artefacts is expressive: Art works can be examined "under microscope and x-ray". Media archaeology here literally means temporal analysis of cultural objects by technical, non-human agents, measuring media by archaemetrics. (Riedere, Von Rohr, 1973)

Archaeology is not just an auxiliary discipline to history, but as well a genuinely alternative model of processing data from the material archives of the past. Radical materialism is a provocation to the historical discourse itself.

While historical discourse strives for narrative coherence, the archaeological aesthetics deals with discrete, serial strings of information which – in an age of computing – gains new plausibility against literary forms of historical imagination developed in the nineteenth century.

In a methodological sense, there is a structural affinity between computer-assisted archaeology as material-orientated science and philology – as long as its hermeneutic method is being replaced by statistical analysis (Boneva, 1971). Inbetween the material monument and the philological record stands the inscription (Stefan, 1971).

The dissonance between archaeology and history is exemplified by the controversial interpretations of an ancient inscription discovered in Italy some thirty years ago, the Lapis Satricanus which seems to bear (and thus authenticate) the name of one of the founders of the Roman republic hitherto considered to be a rather fictitious character in ancient historiography. This case at the same time figures significantly

in the methodical debate between the Arnaldo Momigliano and the Hayden White schools of history. The insistence on archaeological aesthetics, i. e. discrete and non-narrative data analysis in the representation of this fragmented bedrock of evidence, turns out to be a quality of resistance against the national or ideological will for narrative myth-building in history.

In his demand for textual information the historian tends to forget about the materiality of the data carrier. Let us take the early Roman inscription stone excavated in ancient Satricum near Rome as an example: "Once the position of the block with the inscription had been photographically documented and sketched […] this and the two others displaying the same characteristics were transported to the Dutch Institute at Rome for preparation of the publication and to await placement in a museum" (Stibbe, 1980, p. 27) – just what happened to the ancient inscription stone figuring in Mengs' above-mentioned *Allegory of history*.

The act of discursivation, i. e. the scientific publication and subsequent discussion of this archaeological sensation made the material original almost redundant; in the meantime, the inscription stone stayed away hidden in the basement of the Dutch Archaeological Institute in Rome – in the architectural "unconscious", surrounded by metal shelves. As if the hermeneutic focus on the *historical* meaning of the inscription and its communication by publication got rid of its materiality; the material memory here belongs to storage as encryption. History is about discursive reading and writing:

"So often have we all used such publications that we can easily forget that they are *not* the real thing. Even with a scrupulously accurate editor, there are features of the original which are not reproducible: most obviously we lose the document as an archaeological artifact […]." (Dymond, 1974, p. 55)

From epigraphic inscription to volatile data: "Forensic" media archaeology

Matthew Kirschenbaum examines the characteristics that govern writing, inscription, and textual transmission in all media: erasure, variability, repeatability, and endurance (Kirschenbaum, 2008). Significant attention is being paid to storage in its most material incorporation: the hard drive. Understanding storage devices is essential to understand *digital* media. Different from electronic media, which operate predominantly in the "live" mode, digital data and signal processing always require intermediary storage – short time memories which are activated so fast that their sheer existence escapes the comparatively slow human perception.

Towards a material philology: just as the humanities discipline of textual studies examines books as physical objects and traces different variants of texts, "computer forensics" – in a truly media-archaeological way – encourage to perceive new media in both material and logical terms (electronic platforms, programming systems). The *lectio difficilior* in philological hermeneutics is matched here by a close reading of, for example, the function of cheque bits (specially coded additional impulses) in a data word since pulse trains which represent numbers are electronically vulnerable in internal transmission to all kinds of noise which easily leads to a transformation,

thereby: mis-representation of the data word. This is known especially from Compact Discs for music replay where in-built automated correction codes protect against noise in the optical reading of "pits" and "lands" on the disc surface.

Flat temporality: transitive analysis and the microscopic gaze

Does the history of science function as a kind of mental telescope which looks at scientific knowledge from a distance? The metaphor of the micro- or telescope (for identifying ultra- and sub-time-critical intervals) turns history into a laboratory experiment – the artificial extension or condensation of an (apparently) given development. This procedure is well known from the time axis manipulation in the simulation of physical processes like kinematics by analogue electronic computers. There are moments of temporality which depend on such technologies in order to be detected at all – just like Walter Benjamin defined the "optical unconscious" as revealed only by the photographic lens.

The meaning of tradition shifts its focus from its previous emphatic macro-temporal ("history") notion to the analysis of the time-critical and time-basing micro-mechanisms of transmission. While tradition has been associated with long-time diachronic memories across deep historical time so far, this emphatic horizon now shrinks to a mere almost synchronous extension of the present, its re- and protentive short-time memory *online* – a dramatic shift of the temporal prefix in the age of algorithmic, that is: regenerative (instead of inherited) memory.

Literature

Benjamin, W. (1978). *Gesammelte Schriften*, (2nd ed., edited by Rolf Tiedemann / Hermann Schweppenhäuser). Frankfurt am Main: Suhrkamp.

Boneva, Liliana I. (1971). A new approach to a problem of chronological seriation associated with the works of Plato. In F. R. Hodson, D. G. Kendall, and P. Tautu (eds.), *Mathematics in the archaeological and historical sciences : proceedings of the Anglo-Romanian Conference, Mamaia, 1970*. Edinburgh: Edinburgh University Press.

Carpo, Mario (1998). "Descriptio urbis Romae". Ekphrasis geografica e cultura visuale all'alba della rivoluzione tipografica. *Albertiana, 1*, 121-42.

Chun, Wendy (2011). The Enduring Ephemeral, or The Future Is a Memory. In Erkki Huhtamo, Jussi Parikka (eds.), *Media Archaeology. Approaches, Applications, and Implications.* Berkeley, Los Angeles, London: University of California Press.

Cramer, Florian (2002). *Concepts, Notations, Software, Art* (March 23rd, 2002). Retrieved from http://www.netzliteratur.net/cramer/concepts_notations_software_art.html

Dawkins, Richard (1989). *The Selfish Gene*. Oxford: Oxford University Press.

Dymond, D.P. (1974). *Archaeology and History. A plea for reconciliation*. London: Thames and Hudson.

Eliot, T.S. (1949). *Selected Essays*. London: Harcourt Brace. (1st ed. 1932).

Engel, Friedrich Karl (1988). A Hundred Years of Magnetic Sound Recording. *Journal of the Audio Engineering Society, 36*(3), 170-178.

Esch, Arnold (1985). Überlieferungs-Chance und Überliefungs-Zufall als methodisches Problem des Historikers. *Historische Zeitschrift, 240*, 529-570.

Fynsk, Christopher (1992). The Claim of History. *Diacritics, 22*(3-4), 115-126.

Galison, Peter (1997). *Image and Logic. The Material Culture of Microphysics*. Chicago: University of Chicago Press.

Gelernter, David (1997). *Machine Beauty. Elegance and the Heart of Technology*. New York: Basic Books.

Ginouvès, René, & Anne-Marie Guimer-Sorbets (1978). *La Constitution des Donnés en archéologie classique. Recherches et expériences en vue de la préparation de bases de donnés*. Paris: Éditions du CNRS.

Gossman, Lionel (1983). Orpheus Philologus. Bachofen versus Mommsen on the Study of Antiquity. *Transactions of the American Philosophical Society, 73*(5), 1-89.

Gumbrecht, Hans-Ulrich, & K. Ludwig Pfeiffer (eds.) (1988). *Materialität der Kommunikation*. Frankfurt am Main: Suhrkamp.

Heidegger, Martin (1998). Logik als die Frage nach dem Wesen der Sprache [*1934]. In M. Heidegger, *Gesamtausgabe, vol. 38*. Frankfurt am Main: Vittorio Klostermann.

Heyer, Paul, & David Crowley (1995). Introduction. In Harold A. Innis, *The Bias of Communication* [*1951], Toronto, Buffalo, London: University of Toronto Press.

Horatius, *Carmen III, Ode 30*.

Hunger, H. (1971). *Spätbabylonische Texte aus Uruk, vol. I*. Berlin: Gebr. Mann.

Kemp, Wolfgang (ed.) (1980). *Theorie der Fotografie I (1839-1912)*. Munich: Schirmer/Mosel.

Kirschenbaum, Matthew (2008). *Mechanisms. New Media and the Forensic Imagination*. Cambridge, MA: The MIT Press.

Kittler, Friedrich (1989). *Die Nacht der Substanz*. Bern: Benteli.

Krapp, Peter (2011). *Noise Channels. Glitch and Error in Digital Culture*. Minneapolis, London: University of Minnesota Press.

Kusch, Martin (1989). Discursive formations and possible worlds. A reconstruction of Foucault's archaeology. *Science Studies 1*, 17.

Lavers, Annette (rev.) (1985). On: Michel de Certeau, Writing versus Time. *History and Theory, XXII*, 3. (Revised version of (1980) in M.-R. Logan, J. F. Logan (eds.), Rethinking History. Time, Myth, and Writing. New Haven: Yale French Studies 59.)

Lazzarato, Maurizio (2002). *Videophilosophie*. Berlin: b-books.

Lubar, Steven, & W. David Kingery (eds.) (1993). *History from Things. Essays on Material Culture*. Washington, London: Smithsonian Institution Press.

Marker, Hans-Joergen (1998). Data Conservation at a Traditional Data Archive. In Edward Higgs (ed.), *History and Electronic Artefacts*. Oxford: Clarendon Press.

Moberg, C.A. (1971). Archaeological context and mathematical methods. In F. R. Hodson, D. G. Kendall and P. Tautu (eds.), *Mathematics in the Archaeological and Historican Sciences*. Edinburgh, Chicago: Edinburgh University Press / Aldine Atherton.

Nita, Sorin Cristian (1998). Establishing the linkage of different variants of a Romanian chronicle. In Edward Higgs (ed.), *History and Electronic Artefacts.* Oxford: Clarendon Press.

Parikka, Jussi (2012). *What is Media Archaeology?* Cambridge MA: Polity Press.

Polanyi, Michael (1958). *Personal Knowledge*. Chicago: The University of Chicago Press.

Richards, J.D., & N. S. Ryan (eds.) (1985). *Data Processing in Archaeology*. Cambridge (UK): Cambridge University Press.

Richardson, Iain E. (2010). *The H.264 Advanced Video Compression* Standard (2nd ed.). Chichester: John Wiley.

Riederer, Josef, & Alheidis von Rohr (eds.) (1973). *Kunst unter Mikroskop und Sonde. Naturwissenschaftliche Untersuchungen an kulturhistorischen Objekten. Handbuch zur Ausstellung der Staatlichen Museen*. Berlin: Preußischer Kulturbesitz.

Röttgen, Steffi (1980). Das Papyruskabinett von Mengs in der Biblioteca Vaticana. Ein Beitrag zur Idee und Geschichte des Museo Pio-Clementino. *Münchener Jahrbuch der bildenden Kunst, 31*, 189-246.

Sage, Ira M. (1953). Making Machines Remember. *Product Engineering, XXIV*, 141-149.

Shannon, Claude E. (1963 [1948]). The Mathematical Theory of Communication. In Claude E. Shannon and Warren Weaver, *The Mathematical Theory of Communication*. Champaign, IL: University of Illinois Press.

Siegert, B. (2003). *Relais. Geschicke der Literatur als Epoche der Post.* Berlin: Brinkmann & Bose.

Stefan, Alexandra (1971). Applications of mathematical methods to epigraphy. In F. R. Hodson, D. G. Kendall and P. Tautu (eds.) *Mathematics in the Archaeological and Historican Sciences*. Edinburgh, Chicago: Edinburgh University Press / Aldine Atherton.

Stibbe, C. M. (1980). The Archaeological Evidence. In Stibbe et al. (eds.) *Lapis Satricanus. Archaeological, Epigraphical, Linguistic & Historical Aspects of the New Inscription from* Satricum (pp. 21-40). Den Haag: Nederlands Instituut te Rome.

Stockhausen, Karlheinz (1963). Die Einheit der musikalischen Zeit. In Dieter Schnebel (ed.), *Karlheinz Stockhausen. Texte zur elektronischen und instrumentalen Musik, vol. 1*. Cologne: DuMont.

Swade, Doron (1992). Collecting Software: Preserving Information in an Object-Centred Culture. *History and Computing, 4*(3), 227-235

Swade, Doron (1998). Preserving Software in an Object-Centered Culture. In Edward Higgs (ed.), *History and Electronic Artefacts*. Oxford: Clarendon Press.

Warburg, Aby (2010). Mnemonsyne I. Aufzeichnungen (1917-29). In Martin Treml, Sigrid Weigel, Perdita Ladwig (eds.), *Aby Warburg, Werke in einem Band.* Frankfurt am Main: Suhrkamp.

Wiener, Norbert (1948). Time, Communication, and the Nervous System. *Annals of the New York Academy of Sciences, 50*, 197-220.

160

WOLFGANG ERNST
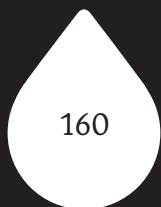
Order by Fluctuation? Classical Archives and Their Audiovisual Counterparts[1]

Technomathematical And Epistemological Options in Navigating Trans-Alphabetical Archives

Introduction: Information practices in media-archaeological reading

Archives today, can be redefined in terms of negentropic systems: the registration of physically real signals by audio-visual media for the "analogue" age, and Digital Signal Processing in our current media practice.

The difference between the symbolic regime of classic, alphabet-based archives, based on alphabetic letters, and audio-visual archives, based on physical signals, is a fundamental one. There are new tools of sound- and image-based search engines, new tools for addressing audiovisual memory, be it cultural, aesthetic, scientific, or otherwise.

For defining what is different in digital archives compared to the traditional ones, arises the issue of how to navigate audiovisually in data avalanches, that is, to find sensual, *aisthesis*-based interfaces for human-machine-memory logistics.

From space-based to time-based archives

While the traditional archive of predominantly textual records provides a spatial order ("l'espace de l'archive", as described by the historian Michel de Certeau), to be transformed into "history" by the act of writing, the audiovisual archive itself takes place in time, beyond the scriptural regime (which is the realm of historiography). We can observe an epistemological dimension: the transformation of the classical, datacarrier based, material storage "archive" into an archive in electronic motion, in electromagnetic ephemerality and latency. The gain of flexibility and computability is paid with a loss of durability.

When a few years ago the Cologne Municipal Archive materially collapsed, it became apparent that most records, though having become dirty and mutilated, materially survived this catastrophe, astonishingly resistant against the pressure of stones. In a similar way the first generation ("analogue") audiovisual storage media turned out to be surprisingly resistant against temporal entropy (like the Edison-cylinder and gramophone records, as well as daguerreotypes, photographic negatives and film on celluloid). More delicate is the destiny of cultural memory based on electromagnetic storage; digital media, finally, tend to divest themselves completely from their material embedding – losing the "touch ground" by becoming technically "virtual".

Traditional physical storage media, whether with spatial or with temporal "bias" (as described by Harold Innis) have been literally inscribed; using writing means, the information to be stored is introduced onto the device. On the contrary, latent storage devices (such as magnetic tape for audio and video) only reveal their memory content in the dynamics of the electromagnetic field ("induced"). Electrotechnical storage media used to take place in a sphere which is different from the scriptural regime of the classical archive – a regime which, on the level of alphanumeric codes, unexpectedly returns in techno-mathematical machines.

The recording of the acoustically or optically physical signal is "more real", as opposed to symbolic notation by the alphabet, not only in a technical but also in an epistemological way: the signal is indexically tied to the physical world, while the culturally coded symbol is arbitrary. With computing, though, this dialectic opposition becomes synthesised, since Digital Signal Processing (notably sampling of audio events) is a function of discrete symbolisation, a re-entry of the "alphabet" in numerical and logical form. Just like the electronic revolution in mass media communication devices like radio and television has resulted in a kind of "secondary orality" (Walter Ong), communication based on the symbolic machine (computing) has resulted in a (hidden) secondary alphabetic revolution, with bits and bytes inheriting the typeset, but different from the printing culture in a dynamic way. This return is a temporal figure which cannot be reduced to the linearity of media history; we are confronted rather with a media-archaeological contemporalisation, a kind of Möbius loop.

With computing binary data, the "archival" symbolical regime returns into audiovisual media themselves, but in a different way: no more signals, but the alphanumerical code. Trans-alphabetical archives as data spaces are not predominantly based on the phonetic alphabet any more (which is the message of the medium archive in the traditional sense – whatever its "content" which is targeted by historiography). This implies a profound mathematisation (instead of narrativisation) of the archive.[2] The archive based on processing binary numbers *can* cope with traditional textual as well as audio and visual records.

---

[2] In French *conter* instead of *raconter*, in German *zählen* instead of *erzählen*.

No more metadating

Traditional, scripture-oriented memory devices demand "a means of tagging the
information so that it can be easily selected when wanted" (Sagem 1953, p. 141),
such as ordering by tree-like *thesauri*. Instead of such metadating there now is the
option of *hashing*, of ordering media from within their genuine media qualities
(f. e. image archives pixel-wise). Whereas the World Wide Web (and the Google
ranking algorithm) reduces hypertextuality to links and statistics between
documents, the original vision developed by Theodor Holm Nelson in 1960
(itself inspired by Vannevar Bush's vision "As We May Think" from 1945) and
finally released in 1998 allows for addressing each single alphanumeric symbol in a
text directly from "abroad", while at the same time keeping a copy of the complete
document at one moment in time in addition to all later variations – a truly
"archival" impulse.

In dynamic data retrieval, it is possible to navigate through large amounts of
photographic and moving images without being guided by verbal language, by
immediate access, unfiltered by words, due to the mode of digital image existence:
the alphanumeric code as the symbolic regime of the digital image. Expressing
pictures by numbers undoes the traditional dichotomy between image and
metadata; both implode into binary numbers.

But on that media-archaeological level, such a two-dimensional set of data is simply
a data format which becomes an "image" only in human cognition, and by verbal
description (*ekphrasis*). Without iconological interpretation of certain visual
patterns (Erwin Panofsky), an image would just be a cluster of data. Optical signals
become information „in the eye of the beholder" only, while the computer can deal
with the symbolical analysis of physical data without the imaginary.

What digital space allows for instead is the option of navigating images in their own
medium – without changing from visual to verbal language. In digital space, the task
of *searching images* does not only mean searching for images, but has a second,
reverse meaning as well: images that can search for similar images, without the
interception of words. Navigating in *Dataland* (as designed in 1973 by William
Donelson), not in the alphabet.

Different from printed letters in a book, the symbols in digital dataland are arranged
and distributed algorithmically – a dynamic that matches the kinetic nature of
orderly movement itself.

How do humans interface to images in a nonverbal way? Let us thus search for visual
knowledge not by metadating images, but within the visual endodata: entering the
image itself (data-immersion).
A critique of the notion of „metadata" draws on the assumption that there is
knowledge already *within* the digitised or born-digital image, a kind of knowledge
which can be grasped below verbal description. Let the image be informative itself –
by means of operating with values that are already intrinsic to the image in a digital
culture when the essence of the image itself dissolves into alphanumerical data.
Algorithms can find, for instance, all edges in a bit-mapped image. What looks like
images, in media-active perception, is rather a function of mathematical data sets.

There is a knowledge already implicit, „dormant" within the electronic images, a kind of compressed virtual knowledge, which – different from external inscriptions (metadata) – waits to be uncovered from within.

Digital databanks of images, when cleverly addressed, render a kind of knowledge which has otherwise been unimaginable in culture. Digital images render aspects of visual knowledge which only the *medium knows*, virtually in the „unconscious" of the data. The media-archaeological program is to uncover such *virtual visual knowledge*.

Any archival record, as opposed from being looked at individually, gets its meaning from a relational structure (which is the archival structure), the relationship to other documents. But opposed to the archival algorithms (its taxonomies), In most media archives, navigation through large amounts of images still requires verbal or alphabetical metadating. To get a videotape from a media library, we still must type a verbal search term into the interface.

Most videotape extraction in film archives has been done by the grip on the whole tape, the storage medium – entering the archive, but not accessing its smallest elements (what the Greeks called *stocheia*, the name for both alphabetic letters and mathematical numbers). The computerisation of such media archives now promises that data objects that traditionally resisted the human attempts to describe them truly analytically will be finally opened – now that images are being understood themselves as data sets, as clusters of grey and colour values.

Addressing and sorting nonscriptural media remains an urgent challenge which, since the arrival of fast-processing computers allowed for digitising analogue audiovisual material. The result is not necessarily better image quality but, rather, the option to address not just images (by frames) but every single picture element, each pixel.

Images and sounds have become calculable and thus capable of being exposed to pattern recognition algorithms. Such procedures will not only media-archaeologically "excavate" but as well *generate* unexpected optical statements and perspectives from an audiovisual archive that can, for the first time, organise itself not just according to metadata but according to its proper criteria – visual memory in its own medium (endogenic).

Contrary to traditional semantic or iconological research in the history of ideas, such an endogenic visual archive will no longer list images and sequences according to their authors, subject, and time and space of recording. Instead, digital image data banks will allow visual sequences to be algorithmically systematised according to genuinely iconic notions and mediatic rather than iconological commonplaces, revealing new insights into their non-symbolical characteristics. A predominantly scripture directed culture still lacks the competence of a genuine *visual dictionary* as made possible for digitally annotated video analysis which allows, e. g., for procedures of dynamic reorganisation." (Ekman, Friesen, 1969)

The real "iconic turn" in addressing photographic images in archives is still to come – a visual sorting on the threshold of digital image processing and retrieval. While visual and acoustic sources contain types of information and aesthetics a text can never convey, the book or the digital text as a verbal research tool have been much easier to handle comparatively than large amounts of images and sounds; that is why the library is still the dominating metaphor of cultural memory. Since calculating and storage capacities of computers have increased significantly, whole audiovisual archives thus become calculable – at least on the level of pixel or scan values. Images and soundtracks can therefore be made accessible in their own medium, with perfectly adequate algorithms of shape and pattern recognition being available. Suddenly, images can be retrieved according to their own properties – that is, not only by the grace of the accompanying text. The mathematician David Mumford (1999) reduced the vocabulary of picture elements in Western visual culture down to twenty-three elements – almost like the letters of the (Greek) alphabet. Image-endogenous systems of classification replace metadating, such as geometric topologies of image or even cinematographic sequences.

Computing thereby offers the possibility of applying non-semantical image sorting programs which create a strictly form-based image assortment – as envisioned by Heinrich Wölfflin in his *Kunstgeschichtliche Grundbegriffe* a century ago. Image-based image retrieval operates in harmony with the mediality of electronic images, for techno-mathematical memory can open up images according to their genuine optical enunciations.

In his film called *Eye / Machine*, Harun Farocki directed attention to *operative images*. So-called intelligent weapons are data-driven by matching images. But visual search engines that can deal with semantic queries are not restricted to military or commercial usage any more, but have become culturally driven in "Digital Humanities". Calculating images, MPEG-7 allows for "layered" image composites and discrete 3D computer generated spaces; according to Lev Manovich the shift is from a "low-level" to "high-level" metadata that describes the structure of a media composition or even its semantics. Digital technologies liberate images from cultural contentism.

For monitoring to process large amounts of electronic images such as human faces, such systems have to get rid of semantic notions of *Gestalt*. The IBM Query By Image Content software did not try to radically decide in the quarrel between semantic *versus* non-semantic information, but rather to distribute the task according to the respective strength in the human-machine interface:

"Humans are much better than computers at extracting semantic descriptions from pictures. Computers, however, are better than humans at measuring properties and retaining these in long-term memory. One of the guiding principles used by QBIC is to let computers do what they do best – quantifiable measurements – and let humans do what they do best – attaching semantic meaning" (Flickner, 1997, p.8).

– which establishes a cybernetic feedback-loop between man and machine, between analogous and digital dataprocessing, thus not trying to efface, but to creatively enhance the human-computer-difference where they meet on the interface.

Archives of temporality

Audiovisual media address humans at the existential essence of our sensation of being. They regenerate temporal experience, thus addressing the human on the sensory (aesthetical, physiological) level as radically present, while our cognition puts it into a "historical" context: here a dissonance takes place, a gap opens.

Furthermore, let us accentuate a clear separation between "audio" and "visual". The one is physical vibration, mechanical impulse, the other refers to the electromagnetic spectrum, a sense organ for "radio" in terms of radiation; ears and eyes are completely different data processors.

The traditional archival regime refers to the symbolic order which is mainly alphabet-based; the audiovisual archive in the age of analogue media technologies to the actual recording of physical signals. The proof lies in the fact that even noise can be registered as audio, to be used for example in radio plays. But even this memory of noise is subjected to the archival regime: there are catalogues and inventories of noise in sound studios and in the archives of broadcasting stations, where a world of noise is ordered according to its alphabetic denomination – subjected to the logo-centristic writing once more.

A new message of New Media: the temporalised archive

Inspired by artistic practice in modernism, media-theoretical analysis in the McLuhan tradition focuses on the message of the medium itself. Applied to memory agencies and especially the "digital archive", this method demands not only a close analysis of its different technology but a new interpretation of its different epistemological and aesthetical dimension as well. While the traditional archival format (spatial order, classification) will in many ways necessarily persist, the new archive is radically temporalised, ephemeral, multisensual, corresponding with a dynamic user culture which is less concerned with records for eternity but with order by fluctuation. New kinds of search engines will not only answer the needs of variable media arts but will develop into a new "art of the archive" itself.

In terms of computer science and communication, the algorithmisation of archival data results in the streaming archive which can be almost immediately accessed online. A problem arises: how can the concept of the archive be opened to "heterochronic" experimentation and at the same time fulfil its traditional task of keeping a well-defined order intact for transmission into future memory? In classic archival terminology, the archival task and the notion of reconfigurability have an oxymoronic relation. For sure, what is new in the so-called digital age, is the "permanent temporality" not only of the archival records but of its archival infrastructure themselves (called hardware and software). So, the traditional "time base" of archive itself becomes a function of temporal change, requiring a differential analysis (in mathematical terms).

Different from the traditional script-based institutional archive, the multimedia archive (as organised by the Internet) becomes radically temporalised. It is rather hypertemporal than hyperspatial, being based on the aesthetic of immediate feedback, recycling and refresh rather than on the ideal of locked away storage for

eternity. The aesthetics of recycling, sampling and cultural jamming is a direct
function of the opening, the openness and the online availability of multimedia
archives.

Two cultures of memory co-exist in our present age: there is still the symbolical,
letter-based regime of the archive and the library which have been defining the
cultural order for such a long time; then a different regime has emerged,
electrotechnical media able to record and to memorise physiologically "real"
gestures, sounds, movements, images, on the sensual and signal processing level. In
Samuel Beckett's one-act drama *Krapp's Last Tape* (1959), immediate recording of
the voice of the protagonist has replaced the traditional written diary. On Krapp's
desktop, a written "archival" inventory and the magnetic tapes co-exist, which
cannot be deciphered directly by human senses but require a true archaeologist of
media memory, the magnetophone itself. Whereas the alphabetic code (the archival
symbolic) depends on being processed by humans in the act of reading, audiovisual
records can only be "deciphered" by media machines such as the phonograph, the
gramophone, the magnetic tape recorder, the video recorder, computing.

The main task of the traditional archive so far has been to keep legally valid
documents intact for proof and re-use. Once the archive is being searched for
different purposes, mainly by historians, this leads to a misreading of its
administrative nature. The aesthetics of the archive is radically non-narrative.
Transforming such records into a historiographical narrative is an act of misreading
the (in)formation of the archive in an effort to humanise it. When the French
historian Jules Michelet visited the parliament archives in Paris to write about the
recent past of the French revolution, he almost believed he could hear the obstinate
murmur of documents, the voices of the dead – as if recorded on gramophone, so to
speak. Romantic historical imagination, in many ways, prefigured the technological
media of later epochs, in between archival phantasms and auditory hallucinations.

By vocalising silent archival records in his reading performances (his *Memory Arena*
series) the media artist Arnold Dreyblatt imbues memory with a diversity of voices.
The "speaking" archive is a hallucinogenic form of memory, resulting from the
cultural-poetic (or rather prosopo-poietic) phantasms of trying to "speak with the
past", as confessed in the introduction of Stephen Greenblatt's *Shakespearean
Negotiations* which became a pamphlet for the method of "new historicism" in
literary studies: "It all began with the desire to speak with the dead" (Greenblatt,
1988, p.1).

The difference between library and archive

Cultural administration uses the term "ALM sector", signifying the trinity of
archives, libraries and museums. But the difference between library and archive is
decisive – a difference which is grounded in its media of support and of logistics. Any
public "Mediathèque" is rather a library than an archive. Collections of radio and
TV broadcast programs rather correspond to what is called a "publication", rather in
correlation with the library than with the archive of records hidden from the public
(which exists on the hidden level, the techno-mathematical regime of the "symbolic
machine" called digital computer).

Etymologically the notion of the library stems from the Greek word for the storage site of "book" rolls (*bibliotheké*). It thus addresses one format and one dimension, storage space rather than time (which is the case for the magnetic "roll": spools and tapes). The library belongs to what Marshall McLuhan called *The Gutenberg Galaxy* (1962), the age of the book and the printing press, as opposed to the modern media age. Since Thomas Edison's invention of the phonograph culture has been enabled to store audiovisual signals directly, physically real (the indexical according to Peircean semiotics), bypassing the translation and abstraction into the symbolical code of the alphabet.

Towards the end of the twentieth century, a radical extension of the traditional, book-oriented task of national libraries took place. The Institut National Audiovisuel in France receives reference copies of every audiovisual medium produced in France since 1995. In Norway the legal deposit act has been extended; it includes at least one copy of any information available to the public, *regardless of medium*. This procedure is automated and is governed by a computer programme. Suddenly, the institution of the library is thrown into the modern technological media age. Algorithmic machines have automated a series of complex operations like "harvesting" the national domain name websites. In terms of running time, the National Library of Norway, after less than twenty years, now holds an entire 180 years of playing time of audiovisual material. The consequences for the *aesthetics* of knowledge result in new methods of research like "Digital Humanities".

Search "within its own medium": Towards content-based audiovisual retrieval

Very often legal paranoia (like the copyright mania) leads to progress in developing technomathematical knowledge; mighty algorithmic tools have been developed for fingerprinting of copyright identification, of locating metadata for media content without metadata annotation. One of the most advanced mass applicable content based search engines for audio data is firmly implemented in the iPod. Listening to a song, the device can be directed to the sonic source with the menu option "Music is being analysed", leading to an almost immediate recognition of the song and the option for (paid) downloads of this very song.

The traditional way of audio or image retrieval used to be the manual annotation of such media with text to create a text-based management system to perform the retrieval. Such a literal transcription of audiovisual evidence into symbolic notation is an asymmetrical transformation, reducing the richness of aesthetic signals to verbal semantic. The alternative way (content-based retrieval systems according to the MPEG-7 standards) is to retrieve audiovisual evidence *in its own media* (that is the aisthetic regime): based on such analysis, it is possible to describe sound or music by its spectral energy distribution, harmonic ratio or fundamental frequency" (Kim, Moreau, Sikora, 2005, p. 2), allowing for a comparative classification of sound categories.

Automatic systems that process audiovisual information allow for search engines to query the descriptions stored in a database. Thus, it is possible to identify identical, similar or dissimilar audiovisual content. As long as such low- or high-level

descriptions are (automatically) extracted from the audiovisual record itself, depicting the variation of properties of audiovisual signals over time or frequency, it makes sense to call the resulting database an "archive" rather than a "library catalogue".

The time domain description by the waveform represents a genuine option of multimedia archives, media-archaeologically revealing characteristics of the original audiovisual signal in its very aesthetic existence: the harmonicity of a signal, its tone or image quality, down to discrete segments such as the pixel itself.

Such a very analytic *iconic turn* makes visual memory mathematically accessible; search engines like QBIC allow for image-based image retrieval by similarity or query by image content. A technical dispositive gains power over the human imaginary, opposite to the classical, paper- and text-based archive as the realm of the symbolic. By far the largest image collection, without saying, is the World Wide Web. In order to efficiently retrieve pictorial data from this database, content-based methods are an attractive alternative to the traditionally used method of manual textual indexing Müller, Wallhoff, Eickeler, Rigoll, 1999, p. 12-1).

Classification by autocorrelation

Speaking to the archive does not achieve a real dialogue with the dead; what we hear is rather the echo of our own voice. Computing now allows to subtract voices from other sound sources by automatic subtraction (folding); "silence detection" itself (the silence of archival space, its absence of voices) is a feature in the current MPEG7 standard for multimedia, especially AudioPower.

The detection of voices is achieved by autocorrelation, i.e. the comparison of a signal with itself when shifted on the time axis. The programming language SuperCollider thereby allows the reconnaissance of *periodic* signals, which is, in other words, phonetic language (where vowels and their formants represent harmonic signals, as opposed to aperiodic consonants), and to separate this from non-harmonic acoustics. To classify the sonosphere surrounding us automatically is a feature of this new classification aesthetics. For the video area, feature extraction (as defined by MPEG7 standard) is already at work, but practically not yet implemented – for both epistemological reasons (the cultural lag of "archiving" practices) and for technological difficulties.

Most of current shot transition detection focuses on detecting simple boundaries: cuts. In most software tools for temporal video segmentation the time-evolving media event is transformed by shot detection (key-frames) into static, storyboard like spatial (rather than temporal) arrangement. Remarkably, "memorisation" here is not based on identification (the identity of the positive image), but on the "kinem", the image-difference, a Cartesian (and de Saussurean) aesthetics of calculation.

Michel Foucault, in his analysis of *The Order of Things*, has elaborated on this epistemological transition between the époque of similarity to the époque of differences.

Content-based image retrieval often yields surprising results in the first series of "similar" images. If we put an image of the Eiffel Tower in Paris into the machine to be matched by similar images, we might get an image of "Big Ben" in London which obviously looks different.[3] But instead of aiming at eliminating such mismatches, f. e. by making the program "learn" the user preferences, as being attempted by most user-orientated commercial image sorting software, we might aesthetically learn from the logic of the computer, looking at an image from a different perspective, asking: what is it that makes an apparent different image "similar" to the reference images from the point of view of info-aesthetics.

An equivalent for such pictorial matching by similarity in the dynamic field is the music finder *mufin* which chooses a cluster of songs in the databank according to the requested moods (respecting tempo, style, instrumentation and so forth). Resulting in findings which have never been searched for, this is genuine "info-aesthetics", according to which the degree of surprise corresponds with the measure of informative quality.

In the Eifel Tower / Big Ben case, there is a similarity in the grid structure of the construction, which becomes apparent once we look structurally, that is: media-archaeologically at the object; 'structure as an analytic tool' and 'structure as the image subject' here coincide).

Apparent (mis)application in automated sound matching as well may open our ears for a different notion of what music is, when the described program "mis"identifies a song with predominant drums and singing as "speech", leading to reconsidering the sonic aspect of (phonetic) language – rather than being a misinterpretation, this computer-based classification reveals the truly mediatic essence of speech (in Indo-European languages at least).

What is informatically at work here is the classification of (micro-)temporal objects (f. e. in spectral analysis), of smallest intervals (by time-discrete sampling), of *delta-ts*. The most relevant spectral components of speech and singing ranges from around 100 to 2000 Hz – which is a temporal, periodic measure.

Fuzzy search

Exploring and developing new options in navigating trans-alphabetical archives depends on two conditions: the technomathematical and an epistemological opening of a different, almost thermodynamical search aesthetics. In his lecture called "The Storm-Cloud of the Nineteenth Century", in 1884 Ruskin answered to the challenge that at his time the library concept of classification by key terms increasingly became substituted by a theory of balance *in motion* (Røssaak, 2010), oriented rather at weather phenomena which bring forces into play that radically alter the traditional order of knowledge: "*order by fluctuation*, a form of order understood as process rather than state", where entropy is not the negation of order but rather its epistemological alternative, "an organizing principle of disorder that only made sense when observed from on high" (Richards, 1993, p. 86f) – just like so-called *distant reading* of big data in Digital Humanities.

When a search engine such as Google is able to predict an influenza, it is because of calculating so-called "swarm intelligence": a growing number of research entries on medicamentation against cold indicates regional agglomeration of disease. Instead of hierarchical classification based on a thesaurus of fixes terms, knowledge is based on statistical probabilities. Data are not being distributed into fixed, unchangeable addresses anymore, but form a cloud.

Classification by correlation: Pockets full of Memories

The media artist George Legrady explores new forms of cultural narratives. In his media-technological installation *Pockets Full of Memories*[4], Legrady – in the best tradition of Bertolt Brecht's "radio theory" from around 1930 – focuses on the potential changing role of the archival reader from receivers to producers, in fact the shift from passive reading to the active archive. In Legrady's installation at the Paris Centre Pompidou, 2001, the audience created an archive by contributing a digitised image of an object in their possession at the time of the exhibition visit. The sum of the archive of objects, organised through a self-organising map algorithm, has been projected on a large gallery wall and the audience will be able to interact, regroup, and reformulate relationships through digital devices, according to both intuitive and classificatory parameters. This true media archive aimed at exploring digital data structures as a site of literally "collective" memory (thus closer to the museum and the library than to the archive in its strict sense).

The archive of objects, once having been converted from analogue (physics) to digital (information), has been stored in a continuously growing database sorted through a complex algorithm and was then projected at a large scale on the walls of the gallery space. The key component, that is: the generative archive of this mechanism is the implementation of the Kohonen Self-Organizing Map (SOM) algorithm that continuously organises the data within a two-dimensional map, positioning objects of similar values near each other to arrive at an overall "ordered" state. This arrival, of course, is permanently being deferred by additional objects: order in fluctuation indeed. Developed by Teuvo Kohonen, a SOM is an algorithm used for representing large high-dimensional data sets. It is an artificial neuronal net capable of adapting to inputs in a self-learning way. The topological model is based on the binary neuronal function which consists of inhibition (hindrance) and coupling. Variations lead to temporary, generative and fuzzy SOMs. So, let us not forget: what looks like iconographically "similar", is in fact a function of mathematical values; similarity is measured by so-called "distance", a numerical parameter.[5]

---

[3] See for example the software tool MoViMos (Modulare Visuelle Mobile Suche), developed by the German Research Center for Artificial Intelligence, Kaiserlautern (www.iupr.org)

[4] Produced in collaboration with Timo Honkela, medialab, University of Helsinki, applying the Kohonen self-organizing map algorithm (SOM)

[5] See for example the tool MoVIMoS for content-based image retrieval, developed by the Forschungsbereich Bildverstehen und Mustererkennung at the Deutsches Forschungszentrum für Künstliche Intelligenz, Saarbrücken (www.dfki.de)

Motion and immobilisation: the audiovisual archive

Whereas the scripture-based classical archive is a static array of records on the grand scale and letters on the microscale, which could be brought in motion only by the act of human reading line by line, the Edison phonograph is the first form of a truly "performative" archive in motion, since its recording (notably the early ethnographic field recordings around 1900, leading to the Vienna Phonograph Archive and the Berlin Phonogramm Archive) is based on a rotating, technically moving apparatus both in recording and in replay (parallel to early cinematographical recording and projection).

In a very simple thought experiment, imagine an early phonographic recording. Whatever the song or speech that will be, parallel to the harmonic timbre of this sound one will for sure acoustically hallucinate the scratching, the aperiodic noise of the recording apparatus, as well. True media-archaeological awareness starts here: the exercise is to be aware that at each given moment media culture is dealing with the past. It is a technological memory. The noise, the scratch of the wax cylinder is the pure message of the medium; in between, the human voice is literally incorporated. Such a recording primarily memorises the noise of the wax cylinder itself – which is not cultural-historical, but cultural-technological, a different kind of information of the real. Media archaeology opens our ears to listen to this as well, not to filter it out. Thereby the phonograph as media artefact does not only carry cultural semantic like words and music, but – like any work of art – is at the same time an archive of cultural engineering as well, by its very material fabrication – a kind of frozen media knowledge, which media-archaeologically waits to be defrosted, liquefied.

Moving Media archaeology: Technology as "archivist" (Phonovision)

For media memory, archival dynamics replaces „archival space" (Michel de Certeau). The earliest known recording from a Television Transmission is the revue *Looking In*, performed by the Paramount Astoria Girls on the BBC Baird television system (30 lines) in April 1933, recorded by an enthusiastic amateur on his recording equipment (the Baird Phonovision system) on aluminium disc. Processed and restored by digital filtering, the key to clarity seems to be movement itself. Any reproduction of one of the 30-line television broadcast as stills in a printing medium (photography in the book), gives a wrong impression of what had been actually seen.[6] Here the time-critical comes in, since printed records (be it texts, be it images) miss a crucial element: the time-base of perception.

"A single frame of the Paramount Astoria Girls may be crudely recognisable, but when seen as a moving dynamic television image, the girls come to life before our eyes. [...] it has much more to do with what we perceive than what is there in pixels, lines and frames. What we are experiencing is not the detail that the eye sees, but the recognition of movement that the brain sees. [...] our brain somehow builds up a model of what we are looking at." (McLean, 2000, p. 211f)

---

[6] See the Restored Video Recordings 1927-1935, *online* http://www.tvdawn.com/recordng.htm

As a physical item and as a technological monument, the Baird *Phonovision*
recording disc is part of the classical archival techniques (subject to inventorisation)
such as any other classical paper record. The difference is *operative*: as a document it
comes only into being, i.e. it becomes "readable", recognisable when being processed
and replayed by a technical medium (first the Phonovision electro-mechanical Baird
equipment, now the digital restoring computer). Furthermore, it needs to be kept
operative by an ongoing medium, which requires the archival artefact to be
processed *online*.[7]

The chrono-archive

Only with the arrival of chrono-photography (Muybridge, Marey) and with
cinematography an impossible occidental dream came true: to catch the dynamic
element in movement, the kinetic. Technical media (both for acoustic and visual
movements) thus have created a new kind of archiv(e-)ability.

It took generations of archivists and librarians to take account of this new option
which for a long time did not fit into traditional archival and classificatory
terminology, which has rather been fixed on static relations between objects (mostly
verbal). The National Library of Australia has been among the institutions which
have created a special task and collection of folk dance and artistic dance (both
scriptural and pictorial, mostly videographical documentation).[8]

Australia had already been path-breaking in institutionalising a multimedia
approach: The National Film and Sound Archive. In the meantime, Europe (where
France has been at the front by establishing the Institut National de l`Audiovisuel,
and recently the Norwegian National Library in Oslo achieved a similar
encompassing task of multimedia and multi-modal archivisation of national
culture) follows Goethe`s idea of a "virtual library": if it is not possible to physically
assemble the audiovisual cultural heritage, there is at least the option of collecting
its information. The Gateway to Archives of Media Art (GAMA) is primarily
dedicated to *ephemeral* forms of art.[9] This ephemerality relates both to the artistic
form (performances) and the techno-electronical content, the so-called "variable
media".

This is the answer to an archival challenge: how can not only material traces and
textual documents, but temporal expressions themselves (movements) be preserved
for future historiographies? Dynamic reiteration of access needs a flexible tool
which allows for the coexistence of different orders without destroying the structure
of the database. Not only the target, but also the mediality of the archive has been
extended.

The answer lies in discovering, reflecting and techno-mathematically realising new
options of flexible access. The most immediate medium for this to take place is the
electronic form of an open source software as content management system which
includes search functions which are not limited to logocentristic addressing any
more.

---

[7] See http://www.tvdawn.com/silvaton
[8] On the National Library of Australia's Collection Development Policy see http://www.nla.gov.au/policy/cdp
[9] See http://www.gama-gateway.eu

Literature

Ekman, Paul, & Wallace V. Friesen (1969). A Tool for the Analysis of Motion Picture Film or Video Tape. *American Psychologist, 24*(3), 240-243.

Flickner, Myron et al. (1997). Query by Image and Video Content: The QBIC System. In Maybury, M. (ed.), *Intelligent Multimodal Information Retrieval*. Cambridge, MA: MIT Press.

Greenblatt, Stephen (1988). *Shakespearean Negotiations: The Circulation of Social Energy in Renaissance England*. Berkeley: University of California Press.

Kim, Hyoung-Gook, Nicolas Moreau, & Thomas Sikora (2005). *MPEG-7 Audio and Beyond. Audio Content Indexing and Retrieval.* Chichester: John Wiley.

McLean, Donald F. (2000). *Restoring Baird's Image*, London: The Institution of Electrical Engineers.

Müller, Stefan, Frank Wallhoff, Stefan Eickeler, & Gerhard Rigoll (1999). Content-based retrieval of digital archives is using statistical object modelling techniques. In *Conference proceedings volume: EVA ´99 (Electronics in the Visual Arts)*. Berlin.

Mumford, David (1999). *The red book on varieties and schemes*. Heidelberg: Springer.

Richards, Thomas (1993). Archive and Entropy. In T. Richards, *The Imperial Archive. Knowledge and the Fantasy of Empire.* London, New York: Verso.

Røssaak, Eivind (ed.) (2010). *The Archive in Motion. New Conceptions of the Archive in Contemporary Thought and New Media Practices*. Oslo: Novus.

Sage, Ira M. (1953). Making Machines Remember. *Product Engineering, XXIV*, 141-149.

176

FIORELLA FOSCARINI AND JUAN ILERBAIG
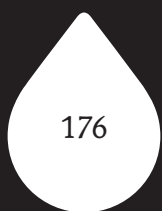
Intertextuality in the Archives

Introduction

Archival science is a contextual science. However, defining what context is and how it can be represented in relation to individual records, archival aggregations, and archives as wholes of records and relationships, continues to be a subject of debate among archival scholars. The notion of context itself, independently of the complexity of the object it characterizes, poses an epistemological dilemma, which literary theorist Jonathan Culler described as follows: "Meaning is context-bound, but context is boundless. … There is no limit in principle to what might be included in a given context … [and] any attempt to codify context can always be grafted onto the context it sought to describe, yielding a new context which escapes the previous formulation" (Culler, 1982, cited in MacNeil, 2004, p. 200).

This contribution begins with a review of various attempts made by different archival schools of thought to frame the "problem of context," from traditional understandings to more recent interpretations of this key concept. It will then focus on the "documentary context," which modern conceptualizations of diplomatics – the centuries-old "science of the diploma" (Duranti, 1989) from which archival science derived – discuss in relation to a specific contextual link among records participating in the same activity, known as "archival bond" (Duranti, 1997). The notion of archival bond, with its characteristics of naturalness, determinateness, necessity, originality, and incrementality, encapsulates the essential properties of a record according to a long-established archival tradition.

By borrowing from other disciplines, such as organizational studies, linguistics and textual studies, the authors will provide insights that point to an expanded and more dynamic view of text-context relationships, a view which better aligns with contemporary archival paradigms invoking constructivist and situated approaches. Rhetorical Genre Studies (RGS), in particular, offers a set of concepts and analytical tools that shed light on the social context of records creation and use, and on the interactions among texts, activities, and agents taking place when we enact records to accomplish our work.

The authors will argue that the notions of intertextuality (Devitt, 1991) and intertext (Christensen, 2016), as reinterpreted by genre scholars following an intellectual tradition that has its roots in early 20th century's semiotics, are especially suited to enrich our understanding of collaborative actions, and the official and unofficial texts that are the outcome and means of such actions. By looking at *intertextual relationships in the archives*, archivists are able to develop an appreciation for the mechanisms involved in the choices made by record creators and users, an appreciation that in turn elucidates context as a situated construct.

Context in archival science

In archives, "the context is all," as Heather MacNeil (1992) entitled her contribution to a collected volume illustrating the fundamental idea of an archival fonds. Even before records become part of complex aggregations (i.e., archives or fonds), in order to exist and have meaning, they must be conceived "in context."

Let us start with some definitions. The Society of American Archivists' (SAA) glossary (Pearce-Moses 2005) defines *record* as "data or information in a fixed form that is created or received in the course of individual or institutional activity and set aside (preserved) as evidence of that activity for future reference." Another definition specifies that a record possesses "content, context, and structure," and that is "used as an extension of human memory or to demonstrate accountability" (p. 326-27).

So, according to the same glossary, context is "one of the three fundamental aspects of a record," (p. 91) more specifically, that part of a record that is not the content (its "intellectual substance" (p. 89)) or the structure (its format or appearance; the way in which the different elements of the record are "organized, interrelated, and displayed" (p. 373)). This latter element of the record, its structure, is somewhat confusingly presented as overlapping the context: it is said to be internal ("relationship of content within a record") as well as external, and the external structure is what "places a record in the context of an order, a series, and a collection."

The SAA glossary offers two definitions for the word *context* in a recordkeeping environment:

> "1. The organizational, functional, and operational circumstances surrounding materials' creation, receipt, storage, or use, and its relationship to other materials.
> 2. The circumstances that a user may bring to a document that influences that user's understanding of the document" (p. 90).

The first meaning is a very complex one. It first divides the context into two kinds of sub-elements: non-documentary circumstances and documentary relationships. The first element is itself subdivided into three kinds of circumstances (respectively related to the organization creating/receiving and using the record, the function in which the record participates, and the operations in which the record is involved). The second meaning may be read as an alternative way of referring to a subset of the first meaning broadly understood; or, from a narrower perspective, it may encapsulate the idea that a record has only one context of creation but is open to being used in different contexts.

It is important to note that the SAA definition of context clearly separates the non-documentary from the documentary context; and it does not analyze in any detail the latter. The organization, its functions, and its operations constitute one part of the context (the "circumstances" of records' creation, etc.), while the "materials" make up a separate part of the context (emerging in the "relationships" to other documents). Similarly, "context" and "materials" are conceived as discrete entities,

the former being broader and more diffused than the latter. It is in other words suggested that the *con-text* surrounds or encloses the *text*, and the two can be analyzed separately.

Let us now consider another perspective on the concept of context, which has more recently been proposed by the InterPARES research project.[1] The first instantiation of this long-running international project, InterPARES 1, draws on "contemporary archival diplomatics" to identify and evaluate the nature of modern records. The project's glossary offers the following definition of *context*: "The framework in which the action in which the record participates takes place" (InterPARES, 2001, p. 2). This definition is action-centered. It articulates the notion of context in relation to the traditional archival science understanding of records as by-products of the actions carried out by individuals and organizations in the usual and ordinary course of their business. The definition continues by partitioning that general context into more specific slices: "The types of context include juridical-administrative context, provenancial context, procedural context, documentary context, and technological context." These different contexts are defined as follows:

- **documentary context**: "The archival fonds to which a record belongs, and its internal structure" (InterPARES 2001, p. 3);
- **juridical-administrative context**: "The legal and organizational system in which the creating body belongs" (p. 5). The normative environment of the records;
- **procedural context**: "The business procedure in the course of which the record is created" (p. 6). Procedure is defined on the same page as "the body of written and unwritten rules governing the conduct of a transaction, or the formal steps undertaken in carrying out a transaction.
  In particular, the legislative machinery set up to carry out a given transaction;"
- **provenancial context**: "The creating body, its mandate, structure, and functions" (p. 6);
- **technological context**: "The characteristics of the technical components of an electronic computing system in which records are created" (p. 7).[2]

The different contexts that a diplomatic approach distinguishes are easy to articulate, similar to unrolling a ball of yarn once you find one of the ends and start pulling from it. First, you determine the creating body for a record or aggregation of records, and this will give you the provenancial context, as you only have to tease out the mandate, structure, and functions of such a body. An examination of the legal and organizational system in which that record-creating body belongs will give you the juridical-administrative context. By studying the laws, regulations, and established practices that regiment the behavior of such an organization and examining the different entities with which it interacts in the performance of its business, one can easily uncover the specific set of actions in which the records at hand were created or received and used. This is the procedural context. This set of

---

[1] For more information on the International Research on Permanent Authentic Records in Electronic Systems (InterPARES) project, see at http://interpares.org/.

[2] It may be interesting to observe that, while the definitions of documentary context and provenancial context together correspond to the first definition of context provided by the SAA, the user perspective insinuated by the second SAA definition of context is not accounted for in the diplomatics-based description of contexts provided by InterPARES 1.

records belongs in a larger whole: the fonds of that organizational body. This is a group of records that, like any other aggregate wholes, possesses an internal structure. The "network of relationships that each record has with the records belonging in the same aggregation" constitutes the so-called "archival bond" (Duranti, 1987, p. 215-216). We will return to this special link among records later in this contribution. For the time being, it will suffice to say that *a record cannot exist without archival bond*. In the absence of this connection to other records participating in the same activity, a record becomes merely a document, information affixed to a medium.

This characterization of context exemplifies the robustness and rigor of the analytical method that is typical of the diplomatic approach. Diplomatics operates by "eliminat[ing] the particularities and anomalies of records in the interest of identifying their common, shared elements" (MacNeil 2004, p. 224). The central idea of diplomatics is "that all records can be analyzed, understood and evaluated in terms of a system of formal elements that are universal in their application and *decontextualized* in nature" (Duranti, 1997, p. 215; emphasis added). According to diplomatics, the form (or internal structure) of a document reveals its context of creation (that is, its external structure); what matters in the context is formally codified, and can be discerned in the document's form by anyone who knows the code.

All contextual elements that do not belong to the specific system of laws, administrative rules, and business procedures, which dictates how actions should be carried out in any given legally-binding situation, are not considered relevant to the understanding of the record from a diplomatic perspective. Furthermore, diplomatics as a system does not capture the open-endedness and lack of linearity of all the contextual elements. These contexts are neatly separated only through an analytical stance that, by abstracting some of the elements of reality, fails to capture the fact that reality is anything but neat.

This approach has profoundly influenced the theory and the practice of recordkeeping. Decontextualization and prescriptiveness are especially the hallmark of most records management literature, which is based on the premise that "[t]he analytical tasks performed by the record professional require that the complexity and messiness of the real world be eliminated, like in a laboratory setting" (Foscarini, 2012, p. 397). As written elsewhere, "[i]n the record disciplines, the world and the word, the context and the text, are conceived as discrete, finite, and dissectible entities. It is part of a record professional's responsibilities to abstract the instantiation of events, which the record impartially encapsulates, from the flux of life, to analyze and describe all elements that participate in the action and the documentation of the action concerned, and to identify and fix those properties that point to the true meaning of the record" (Foscarini, 2015, p. 120).

In recent decades, archival scholars have started to question some of the precepts of traditional archival science by, for instance, investigating its received view of the provenancial context, or context of creation. The traditional, static view of the context of provenance as a single organizational creating body has been reconceptualized by taking into account the dynamic character of the juridical-

administrative and procedural contexts. Organizations have begun to be seen as flexible and adjusting rather than monolithic entities. First, the view of the organization as a rigid administrative structure, best represented in a hierarchical organizational chart, has become untenable since the focus of research has shifted towards the operational processes carrying out the functions of the organization. By prioritizing functional over structural considerations, the very identity of the records creator has come into question (Douglas, 2010; Yeo, 2010b). These changes in our understanding of the context of provenance have come hand in hand with a reconsideration of the traditional interest of archival science in large organizations, often government bureaucracies. Out of this reassessment, archival scholars have become more open to non-bureaucratic environments of business and/or knowledge production (Flinn, 2007; Flinn, 2008; Flinn, Stevens, & Shepherd, 2009).

The new "contemporary archival diplomatics" advanced by the "InterPARES school" has started to be challenged and expanded by bringing in lessons and perspectives from outside the discipline. This process of "situating [diplomatics] within the framework of other disciplinary and philosophical perspectives" (MacNeil, 2004, p. 228) has led to rethink our conceptions of the other contexts of records as well. Archival scholars have long known that the juridical-administrative context and the procedural context can hardly be seen as separate, in that the former continually impinges upon the latter, as changes in the legislation lead to changes in the accepted administrative procedures. Recent explorations of scholarship concerning organizations and organizational culture have provided archivists with new conceptual and methodological tools to reimagine context.

The idea of a "network organization," for instance, with its ad hoc working groups and teams that cross administrative units and professional boundaries for the purpose of working collaboratively on particular projects, is characterized by a flexibility that contrasts with the rigid bureaucracies that reigned supreme during the period of development of modern archival science. Impermanency and fast changing structures and functions have come to characterize the administrative context of records. The procedural context cannot be seen as consisting exclusively of official, written down or agreed upon rules and formalisms. Even in traditional, mono-hierarchical, relatively static bureaucracies, it is often the case that the attitudes and values of individuals or business units in an organization affect the way procedures are executed. Scholars have started to study organizational information cultures, the multilayered complex of attitudes, values, and tacit behavioral norms regarding information that are at work in any given organization and that influence the way records are created, kept, accessed, and used (Oliver, 2008; Oliver & Foscarini, 2013).

The technological context has also been subject to profound reappraisal, in the light of philosophical trends recognizing agency to technology and questioning traditional subject-object positions (Orlikowski, 1992). Some archival scholars have for instance adopted a structurational perspective, which has allowed them to reframe the interrelationship between technology and the structural properties of organizations. In line with this new way of conceiving agent-function-structure relationships, the different layers of context identified within the traditional

archival body of knowledge have started to be seen as interconnected, overlapping, in a continuous state of flux, and inseparable from the object they encapsulate, i.e., the record. Even diachronically considered, context does not change in a linear fashion, with stages in which a realignment of all contexts takes place; instead, there is change taking place at all times, a sort of perpetual contextual motion.

Challenging the documentary context

How about the documentary context? We have discussed how the definition of context in the SAA glossary separated the documentary context from the other surrounding circumstances of the records. The InterPARES understanding of the documentary context identifies it with the traditional layers of arrangement (the archival fonds and its internal structure). This documentary context is seen as the *result* of a series of connections present in each and every record within a fonds. The network of relationships that each record entertains with the records belonging in the same aggregation is known as "archival bond" (Duranti, 1997, p. 215-16). While the SAA definitions of record and context suggest a partial overlapping of both concepts (context is an element of the record, along with content and structure), Duranti sees them as separate: "The archival bond should not be confused with the general tem 'context.' ... [C]ontext is by definition outside the record, even if it conditions its meaning and, in time, its interpretation, while *the archival bond is an essential part of the record*, which would not exist without it" (Duranti, 1997, p. 217, emphasis in original).

The archival bond comes into existence the moment a record is created (i.e., the moment the document becomes a record in connection with other records related to the same function), and is "expression of the development of the activity in which the document participates" (Duranti, 1997, p. 217). Among other characteristics, the archival bond is "incremental" in the sense that it can "grow" beyond its initial connection (in the same way that, to use a biological analogy, a neuron may keep forming connections to other neurons through the life of an individual). However, this "incrementality" is not a notion that traditional archival science has been keen to explore. Since the archival bond is said to find expression in the records classification code, which in turn reflects the functions and activities the records participate in, the underlying idea is that the "originary" moment of records creation, which is "determined" by the function performed by the record in that moment, takes precedence over any future incrementality. The "necessity" inherent in the nature of the archival bond projects a certain "immobility" onto traditional conceptions of the documentary context. The originary context of records creation that the archival bond establishes is in fact dynamic, but only in relation to the "incremental" accumulation of records taking place during the ordinary course of business generating any specific set of records. Following that moment, each linkage among the records that belong to the same activity needs to be fixed in time and space, and must remain stable over time, so as to allow the original context of creation to be knowable.

Among the perspectives that could contribute to revisiting contemporary archival diplomatics, MacNeil mentions "text and discourse analysis," which may help see records as "communicative events and forms of social practice, respectively, and

provide alternative pathways to understanding the nature and purpose of records in a range of record-keeping environments" (MacNeil, 2004, p. 228). Following MacNeil's suggestion, we will now turn to concepts derived from other disciplines, including linguistics and textual studies, with the aim of illustrating some of the limitations of diplomatics and offering new tools for exploring what records are and what they do.

One of the founders of modern linguistics, Ferdinand de Saussure, held a view of language that had intriguing commonalities with the diplomatic perspective of records. Saussure emphasized the relational nature of the linguistic sign, based on the idea that language was a generalized and abstract system, and that the signs in any text were to be understood in their reference to the literary system out of which the text had been produced. Signs lacked independent meaning, in the sense that their meaning was enmeshed in the system of which they were a part. Saussure conceived of the linguistic sign as a "non-unitary" and "relational unit, the understanding of which leads us out into the vast network of relations, of similarity and difference, which constitutes the synchronic system of language" (Allen, 2011, p. 11). These views were adopted by modern literary scholars who looked at texts under a similar systemic light. This systemic view is shared by diplomatics, which sees the record also as non-unitary and relational by definition.

Saussure's perspective was criticized by linguistic scholars who, while following him in accepting the relational nature of the linguistic sign and the literary text, interpreted such relational character as emerging not from the abstractly systemic nature of language but from its existence in "specific social sites, specific social registers and specific moments of utterance and reception" (Allen, 2011, p. 11). In particular, Russian semiotician Mikhail Bakhtin rejected Saussure's "abstract objectivism" for dismissing the social specificity that would give language its very richness. "Linguistics, as Saussure conceives it, cannot have the utterance as its object of study. What constitutes the linguistic element in the utterance are the normatively identical forms of language present in it. Everything else is 'accessory and random'" (cited in Allen, 2011, p. 17). In contrast, Bakhtin noted that what makes words and texts relational is their "addressivity," that is, the quality of always being directed to someone, which only manifests itself in concrete social situations rather than within abstract systems.

Records and context from a genre perspective

Bakhtin's idea of language as a situated, dynamic, and dialogic phenomenon influenced the development of a new stream of scholarship in the area of genre theory known as Rhetorical Genre Studies (RGS) or New Rhetoric. In her article "Genre as Social Action," American communication scholar Carolyn Miller (1984) moved away from the formalistic and abstract understanding of texts characterizing previous approaches, and shifted the focus of genre research to the "recurrent situations" that produce "typified rhetorical actions," or genres (p. 159). RGS is concerned with every-day communicative practices, both written and oral (including all kinds of records, whether organizational or personal, formal or informal), that take place in circumstances that are *recognized* as recurrent by those who attend them (writers and readers, speakers and listeners). This recognition

triggers specific, "typified" answers, which in turn characterize and sustain the social context – made of people, actions, and texts – in which the genres are used. Thus, genres work as "means of orientation," in the sense that they "help us navigate the complex worlds of written communication and symbolic activity" (Andersen, 2008, p. 349) in which we are immersed.

Every community, or group of individuals, who participates in some communicative action in order to get something done, establishes "conventions of discourse" as ways of "acting together" (Miller, 1984, p. 165). Miller (1994) refers to this kind of community, which is dynamic, porous and inclusive (in line with Bakhtin's dialogic view of the world), as a "rhetorical community." Other RGS scholars prefer the expression "discourse community" (Smart 2006). What should be emphasized here is that communities are always culturally and historically situated, and their structure and character are defined by, and at the same time give shape to, the genres that are enacted in such contexts. "In recognizing a text type," Charles Bazerman (2000, p. 16) writes, "we recognize many things about the institutional and social setting, the activities being proposed, the roles available to writer and reader, the motives, ideas, ideology, and expected content of the document, and where this all might fit in our life."

RGS sees genres as helping structure social interaction in the production of work. This structuring function of genres comes from the "bottom," that is, from communities that keep on using "certain material tools [or genres] … in certain ways that worked once and might work again" (Russell, 1997, p. 515). In contrast, the diplomatic view of documents focuses on the "rules of representation" inherent in them, rules that "reflect political, legal, administrative, and economic structures" (Duranti, 1989, p. 15). Diplomatics takes a top-down, normative, or prescriptive approach to the analysis of documents, being concerned with "juridical acts directed to the obtainment of effects recognized and guaranteed by the system" (Duranti, 1989-90, p. 12). This propensity is linked to the historical origins of the discipline and its primary goal of proving the authenticity of documents. On the contrary, RGS is not as concerned with established procedures as it is with the vagaries of process and practice. While one could say that the focus of diplomatics is in the ideal, impeccable form and the fixed, juridical norm, RGS is more interested in the specific and, oftentimes, innovative character of actual genres, their dynamics, and their departure from the norm as a result of their ongoing use. Genres are not defined by their following of strict compositional rules in a top-down fashion, but by the users themselves and their actual, contingent situations of use.

To borrow a metaphor from design, a genre approach to understanding documentary relations and creating information management systems would parallel a view of urban design that, instead of following top-down rules for laying out streets, took a thorough analysis of so-called "cow-paths" as its guide. Cow-paths, also known as "calf-paths" or, less-pejoratively and more poetically, "desire lines," are the paths that pedestrians take informally over a grassy area, rather than using an established, usually paved route. These desire lines are people's chosen ways of navigating space, emerging first as barely noticeable tracks that eventually turn into beaten paths by the recurrence of their use. RGS is chiefly interested in documentary pathways that emerge as desire lines, directly resulting from the social

preferences enacted on a daily basis by the information management practices of workers in an organization, in opposition to the documentary paved roads designed by information engineers and managers. Workarounds are, from this perspective, more important than any official procedure.

MacNeil (2004, p. 230) argued for the need to construct a "'social' diplomatics" that, unlike traditional diplomatics based on the decontextualization of records, would explore the social and cultural contexts of records and achieve a "rich ethnographic description" through its alignment with perspectives from cultural history, historical anthropology, socio-linguistics and semiotics. Instead of dissolving the context into those elements that are encapsulated in the form of the record and those elements that are not and therefore do not matter (de-contextualized away), such a social diplomatics would revel in discovering and making explicit the multitude of connections between the record as text and its complex context.

Recent reinterpretations of the concept of record share the poststructuralist view of language mentioned above. The idea of a record as a "continuum," as an object that is "in a state of always becoming" (McKemmish, 2005) – an idea that since the early 1990s has become part of the archival body of knowledge thanks to the Australian school of recordkeeping – mimics Bakhtin's notion of language as being in a "ceaseless flow of becoming" (cited in Allen, 2011, p. 18). Geoffrey Yeo (2010a, p. 97) referred to speech act theory, with its implication "that each act has a stable and particularized context," to highlight the "performative characteristics" of records. Seeing records as persistent representations of occurrents, Yeo shares with RSG an emphasis on the social agency of records.

In the next section, we will offer a few examples of how RGS scholars have examined the dynamics of genres in ways that aim to suggest an alternative articulation of the documentary context of records.

Genres and intertextuality

An important assumption that RGS shares with archival science is that "no text is single, as texts refer to one another, draw from one another, create the purpose for one another" (Devitt, 1991, p. 336). The study of the interactions taking place among the texts typically produced and reproduced by one particular community, that of tax accountants, allowed RGS scholar Amy Devitt to reveal the social and epistemological characteristics of that community. By borrowing the notion of intertextuality from literary theory, Devitt examined the possible kinds of interactions (among texts, rhetorical situations, agents and purposes) one may observe within a community. The understanding of genres as social action requires that the "genre set" of a community be defined by the participants in that community. Repeated, structured activities and relationships prompt typified responses that draw on previous texts written in response to similar situations. This so-called "generic intertextuality" (p. 338) was implicitly acknowledged by the subjects of Devitt's study. When asked to name the types of texts they typically used to accomplish their work, tax accountants were able to name a few genres that Devitt then categorized as their "genre set."

As mentioned earlier, it is not by looking at laws, regulations, or manuals of procedure that RGS scholars find out what genres are enacted by a specific community. Instead, they apply a bottom-up approach, and while doing so, they often stumble upon "unofficial" work practices or "workarounds" (Spinuzzi, 2003, p. 23). The latter may offer a more comprehensive, colourful, and truthful view of organizational activity than that inscribed in the "official" records. Bakhtin suggested that the continuous emergence of slightly or profoundly transformed genres is the result of the friction between centripetal and centrifugal forces that is inevitable and ongoing in any organization. At the same time, as Paré (2002, p. 60) put it, "genres are socio-rhetorical habits or rituals that 'work,' that get something done," and therefore tend to be produced and reproduced as long as the rhetorical situations remain unchanged. The continuous enactment of the same genres over time helps stabilize those situations, although both the genres and the situations are always "stabilized-for-now or stabilized-enough" (Schryer, 1994, p. 89).

Devitt (1991) identifies a second kind of intertextuality within the tax accountants' texts, which she calls "referential intertextuality," (p. 342) and which has to do with the subject matter of those texts, that is, other texts. Referring to other texts within one's own text is typical of most text-based professions (besides accountants, lawyers, academics, and theologians come to mind), where other texts are explicitly cited, implicitly referred to, or incorporated, as the basis of a writer's authority and expertise.

Finally, the relationship between each accountant's text and those produced previously and subsequently for the same client – a relationship that Devitt (p. 350) labels "functional intertextuality" – is that which contributes to build a "macrotext: the macrotext of that client" (p. 351). Included in this macrotext are the written and oral texts that the community under examination does not produce but receives from the outside. Rather than a genre set, we are now dealing with a "genre system," which Bazerman (1994, p. 97) describes as "interrelated genres that interact with each other in specific settings." Although the purpose, form, and provenance of the genres participating in a genre system may vary, Bazerman adds, "[o]nly a limited range of genres may appropriately follow upon another in particular settings" (p. 98). In other words, functional intertextuality implies that genres do not accumulate randomly, but rather show "some typical sequence (or limited set of acceptable sequences)" (Yates & Orlikowski, 2002, p. 15).

At a level higher than a client's file, Devitt (1991, p. 352) recognizes the existence of "the macrotext of the entire firm's work," corresponding to what archivists would call a fonds. Making a parallel between the notion of macrotext, or genre system, and that of documentary context may appear natural at this point. Both follow the same functional logic and are implicated in purposeful activities. However, as mentioned elsewhere (Foscarini, 2012), the genre system is broader than any file or fonds, as it "reflects a complete interaction including all social relations and the history of the interaction" (Artemeva, 2006, p. 27). Because genre is not only the text, the documentary outcome or residue of a typified activity, but also the context, the recognizable situation that shapes and is shaped by the text, a genre system provides expectations of "what a community does and does not do (purpose), what it does and does not value (content), what different roles members of the

community may or may not play (participants), and the conditions (time, place, form) under which interactions should and should not occur" (Yates & Orlikowski, 2002, p. 18).

More recently, scholars in other areas have used slightly different conceptions of intertextuality to analyze how a variety of interconnected genres participate in the accomplishment of collaborative work in organizational settings such as hospitals. These scholars expand the notions of genre and intertextuality in order to better analyze how records work in specific practical settings, offering situated analyses of the documentary context of records that emphasize the role of texts in constituting practice and handling its contingent nature.

Following Bazerman and others, Carsten Østerlund noted that while the concept of a genre system extends the notion of genre set to all the genres in use, thus instantiating the participation of a plurality of parties in a work process or situation, it does so under the general assumption of a sequential organization of genres. In his study of the texts used in a hospital's emergency room, Østerlund (2007) introduces the notion of "genre combination" to show how genres can be associated or conjoined non-sequentially, by forming accumulations that are driven not by the sequential give-and-take of interaction, but through mere proximity or movement, as forms physically follow the work activities (see also Østerlund 2008).

This idea of genre combinations, Østerlund claims, allows us to better understand how tensions between generic continuity and change, between the stability and instability of genres, can often get resolved not necessarily through the creation of new genres, but simply through the (re)combination of existing ones. As Østerlund (2007, p. 101) contends: "The intertextual readings associated with genre combinations can, in many situations, be established without changing socially recognized expectations associated with an individual genre." Changes in the combinations of existing genres can thus maintain the workability of those genres and allow for the continuing viability of long-established genres even if new situations seem to require generic innovation. Examining these non-sequential genre combinations, Østerlund states, "offers a window into how communities of practice develop 'work-arounds' to buffer themselves from outdated or overly restraining canonical business processes" (p. 105-06).

Another scholar, Lars Rune Christensen, has taken the notion of "intertext," developed by literary theorist Michel Riffaterre, as a way to go beyond the idea of generic intertextuality in the analysis of cooperative work. Christensen (2016) sees the "intertext" as a situational concept, a connection that workers establish between relevant texts in a particular situation, for a particular purpose, and that allows them to "know what to do next." The establishment of intertext "allows us to shift the focus from considering the totality of documents among members of a cooperative work ensemble to considering the perspective of the individual actor making relations between selected texts for a particular purpose" (p. 16). Intertextuality refers to the different ways in which that intertext can be achieved, the different ways of making relations between the elements of a corpus of texts (a genre system) and constructing meaning.

The notion of intertext adds to Devitt's understanding of intertextuality the idea that within a genre system one can distinguish different sub-systems that make sense to each sub-group of participants in a text-mediated interaction. Christensen is arguing that not only does each party have its own set of genres, but also its own situationally-defined sub-system that includes some of the genres from the genre sets of the other participants in the work interaction. However, not every combination is possible, since the genres in a system have affordances, that is, they allow for the establishment of some kinds of intertext and not others, thus encouraging certain kinds of regularities or routines.

In his analysis, Christensen offers another three-way typology of intertextuality. Unlike Devitt's taxonomy, which classifies kinds of intertextuality, Christensen's focuses on ways in which intertextuality can be achieved. He distinguishes between "complementary intertextuality" (when intertextuality is achieved by design, creating forms or documents that complement one another in the information they capture and/or their functionality), "mediated intertextuality" (when intertextual meaning is achieved by means of a third kind of text, such as references to regulatory or legal texts), and "intratextuality" (when the connection is made by juxtaposition or superposition of texts; for instance, by affixing the same identifying label to a series of different forms).

One may argue that these scholars who are using and expanding the views of genres and intertextuality are actually trying to capture the contingency and situationality of practice. In contrast with the traditional archival science's view of the record and its context as fixed and frozen in a moment in time, scholars who adhere to RGS consider genres as "stabilized for now," and regard this imperfect stabilization as depending on the situation and the participants involved. In fact, they have started to delve not only into how genres get combined, recombined, and changed, but also into how they may get combined and recombined so as not to change. They also look at the document as having an in-built portable context or portable place "which helps the reader locate the meaning and the spatio-temporal order out of which it emerges" (Østerlund, 2008, p. 201). That is, the very physicality of the specific time and space in which a document works is yet another element that supports the notion of the inseparability of document and context.

In order to illustrate how some of these ideas may work within a specific archival environment (as opposed to the contemporary organizational environments in which most genre analyses and genre-infused ethnographic work are conducted), let us summarize a case that one of the authors examined in more detail elsewhere (Ilerbaig, forthcoming).[3] The case study analyzes the genres used by Charles Darwin during his years of fieldwork aboard *HMS Beagle* and the years immediately following his trip, during which time he prepared his descriptive work on his collections and developed his first versions of his evolutionary theory.

During the 5-year-long voyage of the *Beagle* around the world, when Darwin was allowed to spend some time on land, his fieldwork consisted of two main activities.

---

[3] Please note that this is a very abridged version of the case study and that a few of the elements of the case have been hypothesized, so the case can illustrate all of the aspects of intertextuality described here. All the materials mentioned here can be accessed through the Darwin Online project (http://darwin-online.org.uk/).

He would observe and write down his observations in a series of small field notebooks he carried with him at all times, and he would collect a variety of animal and plant species, using diverse means, depending on the nature of the organism. Irrespective of the kind of organism, he would immediately affix to it a label with a number. As to the *field notes*, they constituted the first genre of his field set, the starting point of all the other scientific writing he would do. These notes tended to be extremely brief descriptions of a situation, a behavior, or an environment related to the organisms he collected; they were rarely in sentence form, more often just a few words that would later serve as a reminder of the whole situation, and that often would make little sense to anybody else.

At the end of a day or a few days spent collecting, Darwin would sit down and open his *catalogue* or *specimen notebook*, in which he would sequentially record the number from each label next to a name (family, genus, or species) or, more often, brief descriptions of the specimen including the sex, perhaps the locality. He would then use his field notes to bring to memory the events of the day and elaborate on them in a narrative fashion, in a separate notebook. This, his *scientific diary*, is cross-referenced with the catalogue via the specimen numbers. In parallel, Darwin keeps a *personal diary* in which he writes an account of events that in most cases are not strictly scientific. He distinguishes this one from the scientific diary by noting that it contains "not a record of facts but of my thoughts" (Darwin, 1832).

Because Darwin is at the time more an apprentice than a full-fledged naturalist, his collecting activities are oriented towards providing scientific materials for the specialists working in the scientific societies and museums back in Britain. He needs to organize his notes so they can be of use to those specialists, and this is a task he carries out in the long periods when the Beagle is at sea, between continents, or between the continent and an island or group of islands. This task consists in bringing out the collections, sorting them out, rearranging the master catalogues and notes, and splitting them into separate lists and separate sets of notes along taxonomic lines. Unlike all the other genres in this field set, written for Darwin's own eyes, these *taxonomically-arranged lists* and *taxonomically-arranged notes* are intended for the London specialists.

At the end of the voyage, Darwin will spend months, even years, interacting with those specialists, writing scientific descriptions of the different organisms for publication. *Correspondence* with a number of parties will ensue. Specialists, for instance, will inform him of problems and decisions. Darwin will also prod other members of the expedition, trying to collect information that he has not kept (for instance, on the location where some specimens were collected), and they will respond. In some cases, a single note contains Darwin's question, the response (for instance, by Capt. Fitzroy), and Darwin's own notes affixed to the response.

Darwin will also use *scientific publications* from other naturalists, trying to complete his scant knowledge of some areas of the science. The result of this process will be the completion of two kinds of works. The *scientific monographs*, based on the materials he collected and collated in the different texts mentioned above, and dealing with the fauna and flora of the areas visited by Darwin, will be co-authored with the specialists. The other work, the one that will make him famous around the

world, his personal *narrative of the expedition*, will combine materials from the two diaries and form a hybrid sort of scientific and popular work.

All of these genres here described are linked by generic intertextuality. As part of the naturalists' field set, they help structure their fieldwork activities, from the practical constraints of doing science in the field to the social expectations of their counterparts, the specialists in the metropolis. All of these genres are also connected by referential intertextuality, which is established in each and every of the three ways examined by Christensen. First, the number assigned to each specimen collected acts in the way of Christensen's intratextuality, connecting all the sections in the different genres that deal with the same organism and allowing for the formation of intertext. In fact, we could think of the specimen itself as working as a third text for mediated intertextuality, serving as an external normative reference that can be independently read and can help create meaning in a similar way to how regulatory and legal texts do. Finally, each text uses, or is used by, other texts in the set, with paraphrases or quotations allowing us to follow these links. This intertextuality is established almost by design (complementary intertextuality), as the *raison d'être* of each of the different genres is connected with their lying at different points in the several axes that go from the private experience to the public dialogue, from the faraway field to the scientific metropolis, and from the contact with the living organism to its transformation into scientific knowledge.

This brings us to the functional intertextuality at work. The sequentiality inherent in the set of field notes, catalogues, and diaries is due to those "genetic links" between the different genres. Each of them plays a specific role in the process of turning specimens and experiences from the field into natural history knowledge. In this process, they interact with external genres such as the correspondence with specialists and with other members of the expedition, and the publications from other naturalists, and we may consider all of them to form the genre system of Darwin's natural history fieldwork. The role that each one of these genres plays is expressed in its formal characteristics as well as its location along the spatio/temporal and social axes mentioned above. Each text responds to different expectations and has different epistemological affordances. Beyond their sequential combination, they are also differently combined by accumulation in the creation of the final scientific deliverables, as they occupy different "places" in a final axis, the one that goes from the popular to the more technically scientific publications emerging from the voyage.

Conclusion

Genre analysis has been incorporated into organization and management studies (Yates & Orlikowski, 1992), computer supported cooperative work and information systems (Christensen, 2016; Østerlund, 2007; Spinuzzi, 2003), science and technology studies (Bazerman, 1994; Orlikowski, 1992), and knowledge organization (Andersen, 2015), among other research areas. Recently, archival and information science scholars have started to show interest in RGS, some as a set of tools to investigate specific records communities and situations (MacNeil, 2015; McKenzie & Davies, 2012), others as a set of concepts to drawn on in order to bring new insights into one's own disciplinary framework (Foscarini 2012, 2015).

By looking at intertextual relationships in the archives, archivists can develop an appreciation for the mechanisms involved in the choices made by record creators and users, and unpack context as a situated construct. This perspective contrasts with archival science's traditional approach to the documentary context. In this respect, archival science has shown a relative rigidity that could be likened to a view of portraiture that uses lighting, backdrops, and poses in an attempt to capture the personality of a subject or the essence of an activity. The characters are well dressed for the part, central texts or materials are present as symbols of the activity, and perhaps the different phases of the activity are represented in different parts of the composition, as if forming different vignettes. In contrast, like ethnographic studies, the genre perspective takes a more candid approach to photography in which people performing an activity are photographed without their knowledge while going about their daily business, often making-do with ill-suited materials that have outlived their functionality and acting in ways that may not be sanctioned by official procedure or may deviate from the standard-setting norm.

In particular, a number of lessons can be learned from RGS that may help us enhance our archival consideration of the documentary context. First of all, RGS teaches us that context only exists when it is *situated*, as opposed to abstracted or generalized. In a similar vein, records only exist when they are *in use*, that is, actively participating in the production of work, the creation of knowledge, and the construction of social relations and communities. Second, the descriptive approach of RGS calls for a *bottom-up* recognition of genres, instead of their top-down determination, which would be typical of the diplomatic approach. In other words, genres are defined by their creators/users, not by relying on official designations or prescriptive sets of properties. Furthermore, the application of the concept of intertextuality within an RGS framework brings awareness of the *non-sequentiality* (or *relative sequentiality*) of genres. That is, it allows us to see that routines are not established once and forever, but are continuously created, recreated, and transformed through participation in text-mediated interactions. The relative, or imperfect, stabilization of genres depends on both the situation and the participants in it. This *contextual agency of records* is not captured by diplomatics. In addition, RGS's generic and intertextual perspective emphasizes the dynamic nature of records practices (as manifested, for instance, in workarounds) and the emergence of creative combinations of texts that work. Finally, the *dynamism* of intertextual and intratextual relationships, as phenomena that are not guided by functional necessity only, expands our understanding of the archival bond as stabilized-for-now, negotiable, and boundless linkage. The bond among records is now conceived as deriving from fluid, situated interactions among texts, people, and activities, rather than dictated by a predictable set of business rules.

Given these considerations, is it still possible for archivists to distinguish between records and context, to conceive them as discrete entities? Archival science has always been interested in records-in-context, that is, the records and their relationships as inseparable, mutually informing phenomena. RGS pushes the boundaries of this connection by looking at the text (both written and oral) and the context as co-constructing each other within culturally and socially specific situations. Allowing for the inclusion of a more dynamic, dialogic and situated perspective in the archival approach to record-context relationships appears in

alignment with the permanent instability and transformation (in organizational structures, workflows, information technologies, and societal relations) that we are experiencing in our information-driven society. RGS and the notion of intertextuality, with all its nuances, may help develop more flexible, human-centred approaches to records and archives.

Literature

Allen, G. (2011). *Intertextuality* (2nd ed.). New York: Routledge.

Andersen, J. (2008). The concept of genre in information studies. *Annual Review of Information Science and Technology, 42*(1), 339-367.

Andersen, J. (2015). Re-describing knowledge organization: A genre and activity-based view. In J. Andersen (ed.), *Genre theory in information studies* (pp. 13-42) Studies in Information, 11, London: Emerald Group Publishing Limited.

Artemeva, N. (2006). Approaches to learning genres: A bibliographical essay. In N. Artemeva & A. Freedman (Eds.), *Rhetorical genre studies and beyond* (pp. 9-99). Winnipeg, MB: Inkshed Publications.

Bazerman, C. (1994). Systems of genres and the enactment of social intentions. In A. Freedman & P. Medway (Eds.), *Genre and the new rhetoric* (pp. 79-101). London: Taylor & Francis.

Bazerman, C. (2000). Letters and the social grounding of differentiated genres. In D. Barton & N. Hall (Eds.), *Letter writing as a social practice* (pp. 15-29). Amsterdam: John Benjamins.

Chen, Y. (2010). Documenting transitional information in EMR. In J. A. Konstan, E. Chi, & K. Höök (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, April 10–15, Atlanta, USA* (pp. 1787-1796). New York: ACM Press.

Christensen, L.R. (2016). On intertext in chemotherapy: An ethnography of text in medical practice. *Computer Supported Cooperative Work, 25*, 1-38.

Culler, J. (1982). *On deconstruction: Theory and criticism after structuralism.* Ithaca, NY: Cornell University Press.

Darwin, C. (1832). *Letter to Caroline Darwin, Botafogo Bay, 25-26 April, 1832.* Darwin Correspondence Project, Letter no. 166, retrieved from http://www.darwinproject.ac.uk/DCP-LETT-166

Devitt, A. (1991). Intertextuality in tax accounting: Generic, referential, and functional. In C. Bazerman & J. Paradis (Eds.), *Textual dynamic of the professions* (pp. 336-357). Madison, WI: University of Wisconsin Press.

Douglas, J. (2010). Origins: Evolving ideas about the principle of provenance. In T. Eastwood & H. MacNeil (Eds.), *Currents of archival thinking* (pp. 23-43). Santa Barbara, CA: Libraries Unlimited.

Duranti, L. (1997). The archival bond. *Archives and Museum Informatics, 11*, 213-218.

Duranti, L. (1989). Diplomatics: New uses for an old science. *Archivaria, 28*(Summer), 7-27.

Duranti, L. (1989-90). Diplomatics: New uses for an old science (Part II). *Archivaria, 29*(Winter), 4-17.

Flinn, A. (2007). Community histories, community archives: Some opportunities and challenges. *Journal of the Society of Archivists, 28*(2), 151-176

Flinn, A. (2008). Other ways of thinking, other ways of being. Documenting the margins and the transitory: What to preserve, how to collect. In L. Craven (Ed.), *What are archives? Cultural and theoretical perspectives* (pp. 109-128). Aldershot: Ashgate.

Flinn, A., Stevens, M., Shepherd, E. (2009). 'Whose memories, whose archives? Independent community archives, autonomy and the mainstream. *Archival Science, 9*(1-2), 71-86.

Foscarini, F. (2012). Diplomatics and genre theory as complementary approaches. *Archival Science, 12*(4), 389-409.

Foscarini, F. (2014). A genre-based investigation of workplace communities. *Archivaria, 78*, 1-24.

Foscarini, F. (2015). Organizational records as genres: An analysis of the 'documentary reality' of organizations from the perspectives of diplomatics, records management, and rhetorical genre studies. In J. Andersen (ed.), *Genre theory in information studies* (pp. 115-32) Studies in Information 11, London: Emerald Group Publishing Limited.

Ilerbaig, J. (forthcoming). *Instruments of recollection: Field notes, place, and memory in the practice of Darwin's natural history.* Unpublished manuscript.

InterPARES. (2001). *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project*. Retrieved from http://interpares.org/book/index.cfm

MacNeil, H. (1992). The context is all: Describing a fonds and its parts in accordance with the Rules for Archival Description. In T. Eastwood (Ed.), *The archival fonds: From theory to practice* (pp. 195-225). Ottawa: Bureau of Canadian Archivists.

MacNeil, H. (2004). Contemporary archival diplomatics as a method of inquiry: Lessons learned from two research projects. *Archival Science, 4*, 199-232.

MacNeil, H. (2015). The role of calendars in constructing a community of historical workers in the Public Records Office of Great Britain ca. 1850s-1950s. In J. Andersen (ed.), *Genre theory in information studies* (pp. 91-113) Studies in Information, 11, London: Emerald Group Publishing Limited.

McKemmish, S. (2005). Traces: Document, record, archive, archives. In S. McKemmish, M. Piggott, B. Reed, & F. Upward (Eds.), *Archives: Recordkeeping in society* (pp. 1-20). Wagga Wagga, NSW: Charles Sturt University, Centre for Information Studies.

McKenzie, P.J. & Davies, E. (2012). Genre systems and 'keeping track' in everyday life. *Archival Science, 12*(4), 437-60.

Miller, C. (1984). Genre as social action. *Quarterly Journal of Speech, 70*(2), 151-167.

Miller, C. (1994). Rhetorical community: The cultural basis of genre. In A. Freedman & P. Medway (Eds.), *Genre and the new rhetoric* (pp. 57-66). London: Taylor & Francis.

Oliver, G. (2008). Information culture: Exploration of different values and attitudes to information in organisations. *Journal of Documentation, 64*(3), 363-385.

Oliver, G., & Foscarini, F. (2013). *Records management and information culture: Tackling the people problem*. London: Facet.

Orlikowski, W. J. (1992). The duality of technology: Rethinking the concept of technology in organizations." *Organization Science, 3*(3), 398-427.

Østerlund, C. (2007). Genre combinations: A window into dynamic communication practices. *Journal of Management Information Systems, 23*, 81-108.

Østerlund, C. (2008). Documents in place: Demarcating places for collaboration in healthcare settings. *Computer Supported Cooperative Work, 17*, 195-225.

Paré, A. (2002). Genre and identity: Individuals, institutions and ideology. In R. Coe, L. Lingard, & T. Teslenko (Eds.), *The rhetoric and ideology of genre: Strategies for stability and change* (pp. 57-71). Cresskill, NJ: Hampton Press.

Pearce-Moses, R. (2005). *A glossary of archival and records terminology*. Chicago: The Society of American Archivists.

Russell, D. (1997). Rethinking genre in school and society: An Activity Theory Analysis. *Written Communication, 14*(4), 504-54.

Schryer, C.F. (1994). Records as genre. *Written Communication, 10*(2), 200-234.

Smart, G. (2006). Writing the economy: *Activity, genre and technology in the world of banking*. Oakville, CT: Equinox.

Spinuzzi, C. (2003). *Tracing genres through organizations: A sociocultural approach to information design*. Cambridge, MA: MIT Press.

Yates, J., & Orlikowski, W.J. (1992). Genres of organizational communication: A structurational approach to studying communication and media. *Academy of Management Review, 17*(2), 299-326.

Yates, J., & Orlikowski, W.J. (2002). Genre systems: Structuring interaction through communicative norms. *Journal of Business Communication, 39*(1), 13-35.

Yeo, G. (2010a). Representing the act: Records and speech act theory. *Journal of the Society of Archivists, 31* (2), 95-117.

Yeo, G. (2010b). Debates about description. In T. Eastwood & H. MacNeil (Eds.), *Currents of archival thinking* (pp. 89-114). Santa Barbara, CA: Libraries Unlimited.

Zhou, X., Ackerman, M., & Zheng, K. (2011). CPOE workarounds, boundary objects, and assemblages. In J. Dan, R. Olsen, R. B. Arthur, K. Hinckley, M. R. Morris, S. Hudson, & S. Greenberg (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, May 7–12, Vancouver, BC, Canada* (pp. 3353-3362). New York: ACM Press.

196

CHARLES JEURGENS
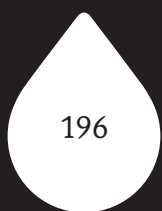
Threats of the data-flood. An accountability perspective in the era of ubiquitous computing.[1]

Overview

In this essay, I argue that ubiquitous computing and the closely related increase in data requires a fundamental reorientation of the recordkeeping community. I explore the effects of data-driven phenomena like big data and smart applications on records and recordkeeping practices from the perspective of its contribution to informational accountability and transparency. I contend that a traditional view of appraisal of recorded data is no longer sufficient to contribute to accountability and transparency. Instead, the focus should be shifted to understanding and managing the assemblages between data and the processing mechanisms (for instance algorithms) in situated practices.

> *There would indeed be no archive desire without the radical finitude, without the possibility of forgetfulness which does not limit itself to repression.*
> *Jacques Derrida*

Introduction

In the mid 1970s, the Italian writer Italo Calvino masterfully depicts the ritual of emptying the trash. In his tale, *La poubelle agréée* he demonstrates the struggle between retaining and discarding. The way people treat their waste reflects the essence of being human, or as Calvino states: "[a]las the unhappy retentive (or the miser) who, fearing to lose something of his own, is unable to separate himself from anything, hoards his faeces and ends up identifying with his own detritus and losing himself in it' (Calvino, 1993, p. 58). Calvino's main character is in a persistent quandary about how to distinguish between the essential and the residue, the meaningful and the meaningless, the relevant and the extraneous. But the perception of what is waste and what is valuable has changed fundamentally in the last few decades. One of the largest European sanitation companies now advertises with the slogan 'waste doesn't exist', since everything can be recycled and reused in the circular economy. This changing perspective bears strong resemblance with one of the core functions the recordkeeping profession is traditionally engaged with: managing abundance by identifying records to be curated and preserved and what

should be discarded. But, analogous to the world of sanitation, the dividing line between valuable information and worthless trash is rapidly blurring.
The recordkeeping community is confronted with this new dilemma since the pervasive recording of data creates unprecedented opportunities in many different domains like health care, crime fighting and societal convenience in smart applications.

Data driven phenomena like Big Data, smart cities and the Internet of things are widely seen as heralds of fundamental societal transformation in a world in which everyone and everything is always connected via information networks. The implications of the computational turn go far beyond the instrumental use of ICT. More fundamental is that the world is increasingly interpreted and explained in terms of data and information. Dutch philosopher Jos de Mul calls it the 'informatisation of our worldview' (De Mul, 2002, p. 130-134). Luciano Floridi designates this turn as the fourth revolution (the three preceding were based on the observations and new paradigms of Copernicus, Darwin and Freud) since human agency in society is entirely determined by ICT which surrounds us. The effect of this informational revolution is, like the previous ones, a fundamental rethinking and repositioning of ourselves into the world (Floridi, 2014, p. 87-94).

The desire to track and monitor nearly everything is not new. States are infamous collectors of information; it is even a prerequisite to possess enough information to be able to create a political space. In 1840 the French politician Pierre-Joseph Proudhon asserted that '[t]o be ruled is to be kept an eye on, inspected, spied on, regulated, indoctrinated, sermonized, listed and checked off, estimated, appraised, censured, ordered about. (…) To be ruled is at every operation, transaction, movement, to be noted, registered, counted, priced, admonished, prevented, reformed, redressed, corrected' (quoted by Scott, 1998, p. 183). What is new are the information and communication technologies that 'record, transmit and, above all, *process* data, increasingly autonomously' and the effect is a strong belief that by doing so society will improve (safer and better quality of life) (Floridi, 2015, p. 52). According to some scholars this makes it viable for governments to record almost everything what people do or say (Villasenor, 2011). Big Data adherents are convinced of the value of data per se and they challenge the necessity of managing information based on the principles of the past. Ralph Losey, an active eDiscovery lawyer foresees that the traditionalist information-management approach based on 'classification, retention, and destruction of information' will be completely superseded within five years. In his view, the 'classify and control lock-down approach of records-management is contrary to the time. Instead of classify and kill, [it is] the *googlesque approach of save and search*'.[2] According to these data hoarders the real efforts to be made are directed towards refining methods of identifying relevant information. Keeping data will become default because as Clay Shirky stated: the problem is 'not information overload. It's filter failure'.[3] The idea of

---

[2] https://e-discoveryteam.com/2015/02/08/information-governance-v-search-the-battle-lines-are-redrawn/?blogsub=confirming#blog_subscription-3 accessed 30 March 2017.

[3] https://www.youtube.com/watch?v=LabqeJEOQyI, accessed 30 March 2017. His vision was disputed by Nicholas Carr, who responded 'It's not information overload. It's filter success' which means that filters push growing amounts of information that is of immediate interest to us, with the result of increasing information overload for individuals, available at <http://www.roughtype.com/?p=1464> accessed at 30 March 2017.

keeping all data, however, is not undisputed. Many scholars envision the future of
information overload in terms of getting stuck in a meaningless data swamp.
Jennifer Gabrys sketches the danger of transforming the archives into sites of digital
rubbish because '[t]he transience and even banality that emerge with electronic
storage extends to new levels, where heartbeats and expiring milk acquire a place as
archive-worthy data. In fact, through the monumental task of archiving everything,
the archive becomes more akin to a disorderly waste site, which then requires
processes of computation to make sense of the welter of material and data' (Gabrys,
2011, p. 120).

In this essay, I explore the implications of this fourth revolution for archival
memory functions in society and more specifically to understand what effects these
data-driven phenomena have on the traditional function of appraisal with regard to
accountability. I will argue that the recordkeeping community needs to put more
effort in rethinking and redefining the prevailing archival concepts and archival
functions. I contend that appraisal remains a meaningful activity in this 'age of
zettabyte' (Floridi, 2014, p. 13), but that the perspective of appraisal in twenty-first
century informational practices is no longer confined to reducing the volume of
records but expanded with the question which components of the constructing layer
of the record are required to keep the quality of records as instruments of
accountability.

Radical turbulences

New technologies that generate, store and transmit data, are changing the nature
of the archive. Geoffrey Batchen writes that the 'archive is no longer a matter of
discrete objects (files, books, art works etc) stored and retrieved in specific places
(…). Now it is also a continuous stream of data, without geography or container,
continuously transmitted and therefore without temporal restriction (…)'
(Batchen, 1998, p. 49; Batchen, 2001, p. 183). The change is not only related to the
abundance of data. Derrida emphasised the importance of understanding the
implications of technologies of communication and recording for the archive. He
coined the term archivisation to express the pivotal impact of the technical means
and methods on what can be archived: 'the technical structure of the *archiving*
archive also determines the structure of the archivable content even in its very
coming into existence.' The performative implications of that notion are far-
reaching, since 'archivization produces as much as it records the event' (Derrida,
1998, p. 17; Manoff 2004, p. 12). In his *Mal d'archive*, which was published in 1995,
Derrida envisaged how for example email will transform the entire public and
private space since '[i]t is not only a technique, in the ordinary and limited sense of
the term: at an unprecedented rhythm, in quasi-instantaneous fashion, this
instrumental possibility of production, of printing, of conservation, and of
destruction of the archive must inevitably be accompanied by juridical and thus
political transformations' (p. 17). The adoption of email in the 1990s is an example
of what Derrida called 'radical and interminable turbulences' (p. 18). New media
transform what can be recorded and archived, and thus what can be used as
evidence. The invention of the camera and phonograph in the nineteenth century
are well known examples of the past. In our time, technologies of Big Data and
Internet of Things cause unprecedented interminable turbulences. In the next

paragraphs, I will first explore the transformative effects of these technologies, then give some examples and I will finish with discussing the implications for recordkeeping concepts and for appraisal and selection.

A need to rethink archival methods

Some leading archival scholars like Frank Upward and Barbara Reed argue that the archives and record profession is facing a widespread crisis. One of the obvious signs of being in crisis is that professionals cannot 'reliably say what a record as a thing is as our conceptual understanding of it blurs into data, documents, information, the archive, and the plurality of archives. The settings in which we manage these converged "things" continues to multiply and increase in complexity. Our new information spaces with their vibrant diversity are paradoxically producing a collapse of collective memory' (Upward et all, 2013, p. 40). There are some parallels to be made with the alarmist view David Bearman already expressed in the late 1980s, when he proclaimed that 'the best methods of the profession were inadequate to the task at hand' (Bearman, 1989, preface). Since Bearman vented his concern, the information-scape has been constantly in transformation. In his time, the late 1980s, the administrative use of Internet was still in its infancy. Tim Berners Lee had just started to work on what would become the world-wide web. Social media were not born yet and the first sms would be sent in 1991. Big Data and the Internet of Things were still a science fiction fantasy. Most of these new media are commonly used nowadays. The computational turn not only affected information and communication behaviour in the personal realm but it profoundly transformed information and communication patterns in administration and business. The computational turn enabled the rise of new economic models which are based on sharing commodities and services, with Airbnb and Uber as the best-known examples. Despite the major changes in the use of ICT, the debate on appraisal and selection has largely remained within the existing document-oriented paradigm. Recently, the Australian Recordkeeping Roundtable paid attention to the implications of the computational turn on recordkeeping functions, including appraisal and selection. Kate Cumming and Anne Picot presented a valuable overview of the challenges appraisal and selection are confronted with. Some of them were diagnosed as technical (new media and applications, networks, changing forms of records, data volumes and storage) and others as organisational (multiple professional responsibilities, decentralised business processes, commercialisation and proprietary systems) (Cumming & Picot, 2014, p. 133-145). They conclude that appraisal in archival institutions is still too much defined as 'a process to preserve a documentary cultural heritage rather than identifying appraisal as laying the basis for practical and accountable recordkeeping'. Although the authors delineate some valuable directions that need to be explored to rethink and reformulate appraisal and call for developing a strategy to prioritise and to employ with business operations, they pay relatively little attention to the fundamental changes that digitisation and informatisation of society have on the attributed function(s) of appraisal. This brings up the following question: what is needed for 'accountable recordkeeping'?

Ubiquitous information technology

In 2011 the authoritative Dutch Scientific Council for Government Policy warned against a precarious lack of awareness among policy-making officials about the far-reaching implications of the networked information structures for the memory functions of iGovernment. The Council emphasised that '[b]oth the importance of 'forgetting' – people should not be judged eternally on the information that government has stored about them – and of saving and archiving require a radical cultural transition and a firmly grounded strategy' (WRR 2011, p. 16 and p. 207). The Council asserted that the government has changed from eGovernment – in which ICT is mainly directed towards providing services – into iGovernment – where ICT changes the relationship between government and citizens because information-flows and data-networks are used for purposes of control and care. The ubiquitous use of memory chips in innumerable applications and functions leads to unprecedented volumes of recorded and processed data. Beyond the three V's (the availability of high volumes, high velocity and high variety of data), it is especially the ability to search, aggregate, and cross-reference large data sets that generate these unprecedented opportunities (Boyd & Crawford, 2012, p. 663). As a result of these innovations, Chris Anderson, editor-in-chief of WIRED magazine, announced the death of theory in 2008 in his much-discussed, contested but nonetheless influential article in *Science* by stating: '(...) faced with massive data, this approach to science – hypothesize, model, test – is becoming obsolete. (...) There is now a better way. Petabytes allow us to say: "Correlation is enough." We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot' (Anderson, 2008). Computer scientist Jim Gray introduced the fourth paradigm of science in 2007. After empiricism (observation and experiment), theory (using models, generalisations, hypotheses) and computation (simulating complex phenomena), science is increasingly based on data intensive computing, which unifies theory, experiment and simulation (Hey cs, 2009). This mixing up of correlation and causality and this naïve belief in the power and possibilities of data to solve present-day problems is typical for these big data adherents.

Data in itself might be seen as innocent, but the processing is definitely not (Rouvroy & Berns, 2013). It is the processing activity that makes data meaningful and transforms data into information. Transforming data into meaningful information cannot exist without a selective perspective. The terms data and information are often improperly used as synonyms. Liebenau and Backhouse make a clear distinction between data and information by defining data as 'symbolic surrogates which are generally agreed upon to represent people, objects, events and concepts' while information is 'the result of modelling, formatting, organising, or converting data in a way that increases the level of knowledge for its recipient', or as they summarise: 'information is data arranged in a meaningful way for some perceived purpose' (Canhoto & Backhouse, 2008, p. 48). The techniques used for modelling and organising data are increasingly computational algorithms. Algorithms are basically a set of rules or instructions to perform a certain assignment in order to process input into output. In the words of the Norwegian media scholar Eivind Røssaak, computational algorithms have become the new lingua franca of codes in the informational infrastructure and they increasingly

rule society and our lives (Røssaak, 2016, p. 34). Compared to human processing, computational algorithms have many advantages since they are much faster, can deal with more complexity and are more accurate than humans will ever be. The downside of this computational processing is that these systems rely on processes and abilities 'that are radically beyond what is possible for human beings to understand' (Danaher, 2016, p. 247). They are black boxes and that is what gives rise to many concerns, because we do not understand how these algorithms operate as the new power brokers in society (Diakopoulos, 2014, p. 2). Critics like Evgeny Morozov and Cathy O'Neil stress that algorithms are constructed models, based on choices what to include and what to leave out. And these choices 'are not just about logistics, profits and efficiency. They are fundamental moral' (O'Neil, 2016, p. 218; Morozov, 2014, p. 182-186). John Danaher warns that the increasingly reliance on algorithms in decision making processes might turn society in an 'algocracy', a governance system in which computer-programmed algorithms are used 'to collect, collate and organise the data upon which decisions are typically made, and to assist in how data is processed and communicated through the relevant governance system' (Danaher, 2016, p. 247). While in a bureaucracy laws and regulations structure and enforce how humans act, in an 'algocracy' the algorithms are the structuring and constraining components. Janssen and Kuk emphasise that algorithms do not work on their own, but form an 'algorithmic materiality', which means that there is an intricate relationality between algorithm, systems, data and humans resulting in a dynamic and 'complex socio-technical ensemble of people, technologies, code developers and designers' (Janssen & Kuk, 2016, p. 274-275), which is very similar to the archive as 'the apparatus through which we map the everyday' (Giannachi, 2016, p. xv).

A few societal examples discussed

There are good reasons to worry about this emerging 'algorithmic governmentality' as some scholars label this data-driven exercise of power and policymaking (Thomas & Berns, 2013; Rodrigues, 2016). Before discussing the archival implications of these socio-technical developments, I want to review some examples. Real time processing of large quantities of data from criminal records, police databases and surveillance data to predict where criminal activities are likely to happen (predictive policing) has already been put into practice in several countries (Joh, 2015). Even in the courtroom computational algorithmic support has been introduced to underpin court decisions. The independent non-profit organisation of investigative journalism *ProPublica*, recently published a series of critical articles on the accurateness of algorithms used in courtrooms to assess the likelihood of recidivism of defendants. *ProPublica* journalists analysed the accuracy of a widely-used risk assessment tool named COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) by investigating 10,000 criminal defendants in Florida and compared their predicted recidivism rates with the actual rates. The researchers found out 'that black defendants were far more likely than white defendants to be incorrectly judged to be at a higher risk of recidivism, while white defendants were more likely than black defendants to be incorrectly flagged as low risk'.[4] Although the Supreme Court of Wisconsin expressed its concern about this race correlation in COMPAS, in an appeal from an order of a circuit court it judged that the evidence-based risk assessment tool COMPAS can be used at sentencing.[5] In the explanation

of its decision, the Supreme Court circumscribed its use by stressing that risk scores are 'not intended to determine the severity of the sentence or whether an offender is incarcerated' and that risk scores 'may not be considered as the determinative factor in deciding whether the offender can be supervised safely and effectively in the community'. Although the Supreme Court agreed that the defendant-appellant was not able to review and challenge how the COMPAS algorithm – which is part of the trade secret of the developer Northpointe Inc. – calculates risk, the order judged the ability to review and challenge the resulting risk scores as satisfactory.[6] Interestingly, one of the judges, Shirley S. Abrahamson wrote a separate consideration in which she emphasised the relevance of recording the use of risk assessment tools. Precisely because scholars were critical on using these risk assessment tools in sentencing, courts should 'evaluate on the records the strengths, weaknesses, and relevance to the individualized sentence being rendered of the evidence based tool (or, more precisely, the research-based or data-based tool)'. Abrahamson recognised that this might be an extra demand on and administrative burden for the circuit courts, 'but making a record, including a record explaining consideration of the evidence based tools and the limitations and strengths thereof, is part of the long-standing basic requirement that a circuit court explain its exercise of discretion at sentencing'.[7]

We need to question how the record is defined if the judges accept that algorithms can be used in sentencing although the algorithm itself, the lens through which the data are filtered, sorted etc., remains a black box because of the mentioned trade secret. The record-making as defined by the Supreme Court has to do with accountability of how the judges use the tools in the process of sentencing, not with the processing activity of the algorithms themselves. This appeal clearly shows the limitations of the traditional scope of the concept of the record. If the informational algorithm remains a closed black-box in cases with far-reaching consequences for citizens (even if the outcomes can only be used as additional information for decisions) the claim that records provide the best means for warranting accountability is severely affected. The ever-increasing interrelationship between man and technology requires a clearer notion of the scope of the record, especially when a relation is made to accountability of decision-making. There are good reasons to redefine the scope of the record in that tight relationship between humans and machines. I agree with Amelia Acker, who argues that examining the infrastructure of records, 'archivists can think big enough about the "black box" and all the layers of construction behind digital records and emerging documentation practices' (Acker, 2016, p. 294-295). One of these layers of construction are the algorithms that are used in the processing of data. The use of 'black-box' algorithms in decision-making processes will be mirrored in the records that are created, and it is not without consequence to the attributed quality of the records as means of accountability. It is imaginable that for some decision-making processes (which immediately shows an additional selection perspective) open and understandable algorithms are required. That is for instance the motive of a motion for the European

---

[4] https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm accessed at 30
    March 2017

[5] Supreme Court of Wisconsin, State of Wisconsin versus Eric L. Loomis on certification from the court of
    appeals, 13 July 2016, available at <https://www.wicourts.gov/sc/opinion/DisplayDocument.
    pdf?content=pdf&seqNo=171690> accessed at 30 March 2017.

[6] Ibid., par. 53.

[7] Ibid., par. 141

Parliament Resolution on Robotics, debated in 2017. Article 12 of this resolution says that it should always be possible to supply the rationale behind any decision taken with the aid of Artificial Intelligence (AI) that can have a substantive impact on one or more persons' lives. It must always be possible to reduce the AI system´s computations to a form comprehensible by humans. Interestingly that same article articulates the necessity that 'advanced robots should be equipped with a 'black box' which records data on every transaction carried out by the machine, *including* (my italics CJ) the logic that contributed to its decisions'.[8] This is in line with the regulation *on the protection of natural persons with regard to the processing of personal data and on the free movement of such data* which was adopted in April 2016 by the European Union. This regulation sets rules and requirements for data-driven automated processing. Every person has 'the right not to be subject to a decision (…) which is based solely on automated processing (…)'.[9] Examples that are explicitly mentioned are automatic refusal of an online credit application and e-recruiting practices without human intervention. Predictive profiling, one of the most fundamental intrusions in a private life which is increasingly used by authorities in fighting crime and terror remains allowed 'where expressly authorised by Union or Member State law'.

Big data analysis is useful for revealing general patterns, but, and that is not always kept in mind, there is always a probability of a mismatch between general patterns and a specific situation (WRR, 2016, p. 27). Scholars, journalists, advisory and legislative bodies warn against excessive techno-dependency and techno-optimism.

The Dutch investigative journalist Dimitri Tokmetzis criticises the naïve way rules are formulated and used in algorithms without paying enough attention to the validity of the underlying assumptions. To give an example: Dutch government assumes that poverty is a risk factor for the education of children. It is possible to design an algorithm to find evidence of poverty in the electronic child records and to make a list of families that need to be watched closely to be able to intervene if necessary. But do we know whether the assumption behind the rule is valid? Is the assumption that defines the rule based on thorough scientific research? (Tokmetzis, 2012, p. 59-60). Scott Mason, researcher at Keele University, also warns against the often-careless way how bureaucrats and policy-makers interpret and contextualise the results of Big Data itself without consulting domain experts to assess the validity of correlations. He is very critical about the claim that Big Data analysis creates the possibility for 'neutral' evidence based policy-making. According to Mason, 'the vast quantities of correlations generated by Big Data analytics act simply to broaden the range of 'evidence from which politicians can chose to support their arguments' (Mason, 2016). In 2016, the Dutch Scientific Council for Government Policy notified a highly undesirable tendency of policymakers to accept the revealed patterns without questioning the validity of the results for specific situations (WRR, 2016).

Archival implications

The aforementioned examples show that computational algorithms increasingly become an integrated part of government processes and decision making. Legal scholars have argued for more than 20 years in favor of more transparency in

automated processing (Kroll, 2015, p. 6). Since records which are created in the course of business, 'provide evidence of actions, decisions, and intentions, both legal and illegal, proper and improper, and wise and misguided' (Cox & Wallace, 2002, p. 4), availability of records is vital for accountability and transparency. Also in the recordkeeping realm, algorithmic tooling is used to manage the growing number of documents and to find relevant information for a specific purpose. The most advanced developments of algorithmic computation in the recordkeeping sphere can be found in eDiscovery and information retrieval applications.[10]

Since archivists claim to play a pivotal role in defending institutional and societal transparency and accountability (Jimerson, 2009, p. 246-252), there is an urgent need for archivists to ruminate what it means to take this role in the era of ubiquitous computing. If records, archives and archivists want to continue to be key players in ensuring and defending accountability, this evokes the question what meaningful recordkeeping is in this new context of data-ubiquity, and at the same time what meaningful records are.

As Upward and others have put forward, this is exactly one of the main challenges the archival and recordkeeping community is confronted with: to clarify how the conceptual relationship between data, records and archives is designated in the era of ubiquitous computing. The traditional record was based on fixity and stability in a material sense. What has fundamentally changed is the possibility to produce different aggregates out of the same recorded data, which means, as Bruno Latour (2009) writes, 'that the whole has lost its privileged status' which makes us aware of the fact that the whole is always simpler than the parts (p. 198). The written record used to have the shape of an entity (the whole) in which the parts (words, sentences, paper, lay out, signature etc.) were a fixed materialised aggregate. Since the computational turn, it is possible to use the same recorded parts (data) in different configurations simultaneously. The stable whole has been replaced by a 'continually evolving liquid assemblage of action' (Introna, 2016, p. 19). It is as if we construct different types of houses with the same bricks at the same time.

This (informational) fluidity is an important feature of what was designated by Deleuze as an assemblage. An assemblage is in the words of Deleuze 'a multiplicity which is made up of many heterogeneous terms and which establishes liaisons, relations between them (…). [T]he assemblage's only unity is that of co-functioning' (Deleuze & Parnet, 2007, p. 69). In an assemblage, an element can be dissociated from a specific assemblage and continue to function in another assemblage. Assemblages exist merely because of the relationships between the elements. Deleuze emphasises that an assemblage is never technological: '[t]ools always presuppose a machine, and the machine is always social before being technical.

There is always a social machine which selects or assigns the technical elements used' (Deleuze & Parnet, 2007, p. 70). This is an important notion which might be helpful to disentangle the sometimes-confusing relationship between data and records.

In a world of ubiquitous computing the ability to define data-points and to monitor and record data has become infinite. CISCO expects that in 2020 more than 50 billion devices are connected to the Internet and these devices 'require minimal human intervention to generate, exchange and consume data' (Rose cs, 2015, p. 17). In the past, it was a time-consuming human activity to select the elements worthwhile to be recorded. That process, the 'conscious or unconscious choice (determined by social and cultural factors) to consider something worth archiving' was coined by Eric Ketelaar as archivalisation (Ketelaar, 1999). Archivalisation precedes archiving and to understand this process, we need to understand what Hofstede called the 'software of the mind' which is programmed by social and cultural factors and comes very close to Deleuze's 'social machine'. Nowadays all particles that are 'observed' by a machine are recorded, although, and that is not unimportant, the data points to be monitored and recorded still need to be defined and programmed. The recorded raw data in itself is meaningless; these are signals without real significance. Data are only meaningful in relationship with other data, processed in a specific situation. Only from that perspective the concept of the record or archive is a meaningful construct. It means that archives should be seen as Foucauldian apparatuses (of governance), dispositives, machineries of seeing, but, and that has to be emphasised, machineries of seeing from a particular point of view (Giannachi, 2016, p. xv-xvii; McQuillan, 2016, p. 8). Thinking about the archive via the apparatus means focusing on the networked arrangement of media, mechanisms of communication and data processing (Packer, 2010). The archive is meaningless without understanding the interdependency of these socio-technical components and humans. Only then we will be able to understand, as Geoffrey Bowker writes, that every act of permitting 'data into the archive is simultaneously an act of occluding other ways of being, other realities. The archive cannot in principle contain the world in small; its very finitude means that most slices of reality are *not* represented' (Bowker, 2014, p. 1797, italics CJ). One could argue that, compared to the past practices of recording, the number of potential witnesses within a situated practice have incredibly increased by the explosion of sensors and data-points. Nevertheless, what is represented by records is in the end based on situated needs, that define the technical arrangements.

Back to appraisal

I opened this article with Calvino's quandary how to distinguish between the meaningful and meaningless. I conclude with the proposition that in our time of ongoing datafication of society, archivists need to redefine the record and as a consequence of it the recordkeeping mechanism of distinguishing between the essential and residue. I showed that protecting informational accountability requires rethinking the components of the record or archive. The apparatus-view, in which the archive functions as a machine of governance, is helpful to understand the intricate, assemblage-based relationality between the components of the archive. The recorded data is only one element of that machine. What data is relevant, and

which other components are required to create a meaningful record, is defined by the situated context of operation. The Volkswagen emission scandal of 2015 may serve as an example. Advanced software in the diesel engines of Volkswagen could detect when the car was tested and subsequently adapt its emissions during the artificial test circumstances to acceptable levels. Back on the road, the vehicles switched to normal mode with much higher emission rates. It is an example that shows that the recorded data of the tests can only be understood in combination with the software, and that the software can only be understood if the logic of the algorithms is known. If the record only provides the recorded data, it is not sufficient for informational accountability. Above all, algorithms are models written with a specific purpose and they are not neutral nor objective. Understanding the results of algorithmic processing requires at least knowledge of the underlying assumptions of the model and the data which are used by the algorithms. The familiar principle of 'the context is all' is also applicable in this layer of construction of the record.

This has major implications for the issue of appraisal, which gets a much wider scope than just answering the question of keeping or discarding recorded data. From a recordkeeping perspective, the issue is not about data; it is about what people, institutions and communities want to be able to reconstruct for purposes of business, evidence, accountability and memory. That perspective is decisive for answering the question which components of 'the archive as an apparatus' should be preserved in coherence. Providing robust accountability is not an easy and especially not a pure technical task to accomplish. Joshua Kroll, who developed a general framework for accountable algorithms in automated decision-making processes, stresses that accountability requires the possibility to verify that 'the social, legal and political structures that govern an automated process function as they are intended to function' (Kroll, 2015, p. 210). Robust accountability requires involvement in the system design and computer systems should be designed in a way that they are reviewable (Kroll, 2015, p. 188-202). Cathy O'Neil started a business to audit algorithms. In an interview with the Los Angeles Times, she explains 'I don't want to just audit a specific algorithm by itself, I want to audit the algorithm in the context of where it's being used. And compare it to that same context without the algorithm' (Los Angeles Times, 2016). Nicholas Diakopoulos argues that a new accountability perspective is necessary in freedom of information requests. Although there are some examples of successful use of Freedom of Information Act requests to compel disclosure of source codes (Diakopoulos, 2016, p. 59), this is definitely not sufficient to guarantee accountability. He suggests reconsidering FOIA along the lines of Freedom of Information Processing Act which is not so much based on disclosing codes, but allow to submit benchmark datasets the government agency is required to process through its algorithms (Diakopoulos, 2016, p. 59). These are just some examples of efforts to accomplish informational accountability.

Does this imply a profound reorientation of the archival community? Yes and no. No, since it is all about understanding the context. But the efforts to be made to understand the context of creation and use require a reconsideration of the components of the record. The archival community needs to rethink and reconceptualise the essence of a record in a world in which data is ubiquitous, fluid and too abundant to manage and control. If the archival community wants to continue to play a meaningful role in defending informational accountability and

transparency (which includes the historical perspective), a more situated approach is required. Understanding the quality of data and the processing mechanisms of data in situated practices is a prerequisite to be able to play that role. I argue that the apparatus perspective provides a useful framework to understand the archive in situated settings. Only if archivists develop the competences to understand the data assemblages and processing mechanisms in situated practices it will be possible to distinguish between the essential and the residue, the meaningful and the meaningless, the relevant and the extraneous.

Literature

Acker, Amelia (2016). When is a record? A research framework for locating electronic records in infrastructure. In Anne J. Gilliland, Sue McKemmish and Andrew J. Lau (eds.), *Research in the Archival Multiverse* (pp. 288-323). Clayton, VIC: Monash University Publishing. http://dx.doi.org/10.26530/OAPEN_628143

Anderson, Chris (2008). The end of theory: the data deluge makes the scientific method obsolete. *Science*. Retrieved from http://www.wired.com/2008/06/pb-theory/

Batchen, Geoffrey (1998). The art of archiving. In Ingrid Schaffner (ed.), *Deep Storage: Collecting, Storing and Archiving in Art.* Munich and New York: Prestel Verlag and Siemens Kulturprogramm.

Batchen, Geoffrey (2001). *Each Wild Idea. Writing, Photography, History*. Cambridge, MA: MIT Press.

Bearman, David (1989). Archival Methods. In *Archives and Museum Informatics Technical Report #9* (preface). Pittsburgh: Archives and Museum Informatics. Retrieved from http://www.archimuse.com/publishing/archival_methods/#intro

Bowker, Geoffrey C. (2014). The Theory/Data Thing. Commentary. *International Journal of Communication, 8*, 1795-1799.

Cox, Richard J., & David A. Wallace (eds.) (2002). *Archives and the Public Good. Accountability and Records in Modern Society*. Westport-London: Quorum Books.

Boyd, Danah, & Kate Crawford (2012). Critical questions for big data. *Information, Communication & Society, 15*(5), 662-679.

Canhoto, A.I., & Backhouse, J. (2008). General Description of the Process of Behavioural Profiling. In M. Hildebrandt and S Gutwirth (eds.), *Profiling the European Citizen: Cross-disciplinary perspective* (pp 47-63). Dordrecht: Springer.

Calvino, Italo (1993). La poubelle agréée. In I. Calvino, *The Road to San Giovanni.* New York: Vintage Books.

Cumming, Kate, & Anne Picot (2014). Reinventing appraisal. *Archives and Manuscripts, 42*, 133-145.

Danaher, John (2016). The threat of algocracy: reality, resistance and accommodation. *Philosophy & Technology, 29*, 245-268.

Datta, Anupam, Shyak Sen, & Yair Zick (2016). Algorithmic Transparancy via Quantitive Input Influence: Theory and Experiments with Learning Systems. In *IEEE Symposium on Security and Privacy* (pp. 598-617). Retrieved from http://www.ieee-security.org/TC/SP2016/papers/0824a598.pdf

Deleuze, Gilles, & Claire Parnet (2007). *Dialogues II.* New York: Columbia University Press.

Derrida, Jacques (1996). *Archive Fever. A Freudian Impression*. Chicago / London: University of Chicago Press.

Diakopoulos, Nicholas (2014). *Algorithmic Accountability Reporting: On the Investigation of Black Boxes*. New York: Tow Center for Digital Journalism. Retrieved from https://towcenter. org/research/algorithmic-accountability-on-the-investigation-of-black-boxes-2/

Diakopoulos, Nicholas (2016). Accountability in Algorithmic Decision Making. *Communications of the ACM, 59*(2), 56-62.

Diebold, Francis X. (2012). *A personal perspective on the origin(s) and development of "Big Data": The Phenomenon, the Term and the Discipline*. Retrieved from http://www.ssc.upenn.edu/~fdiebold/papers/paper112/Diebold_Big_Data.pdf

Floridi, L. (2014). *The 4th Revolution. How the infosphere is reshaping human reality*. Oxford: Oxford University Press.

Floridi, L. (ed.) (2015). *The Onlife Manifesto. Being Human in a Hyperconnected Era*. Heidelberg: Springer.

Gabrys, Jennifer (2011). *Digital Rubbish: A natural history of electronics*. Ann Arbor, MI: University of Michigan Press. http://dx.doi.org/10.3998/dcbooks.9380304.0001.001

Giannachi, Gabriella (2016). *Archive Everything. Mapping the Everyday. Cambridge*, MA: MIT Press.

Hey, Tony, Stewart Tansley, & Kristin Tolle (2009). *The Fourth Paradigm. Data-intensive Scientific Discovery*. Redmond: Microsoft Research. Retrieved from https://www.microsoft. com/en-us/research/wp-content/uploads/2009/10/Fourth_Paradigm.pdf

Introna, Lukas D. (2016). Algorithms, governance and governmentality: on governing Academic Writing. *Science, Technology & Human Values, 41*(1) 17-49. https://doi.org/10.1177/0162243915608948

Janssen, Marijn, & George Kuk (2016). The challenges and limits of big data algorithms in technocratic governance. *Government Information Quarterly, 33*, 371-377.

Jimerson, Randall C. (2009). *Archives and Power. Memory, Accountability and Social Justice*. Chicago: Society of American Archivists.

Joh, Elizabeth E. (2015). Policing by numbers: Big Data and the fourth amendment. *Washington Law Review, 89*(35), 35-68.

Ketelaar, Eric (1999). Archivalisation and archiving. *Archives and Manuscripts, 27*, 54-61.

Knobel, Cory Philip (2010). *Ontic Occlusion and Exposure in Sociotechnical Systems*. Dissertation at the University of Michigan. Retrieved from https://oatd.org/oatd/record?record=handle%5C%3A2027.42%5C%2F78763

Kroll, Joshua Alexander (2015). *Accountable Algorithms*. Dissertation at the Princeton University. Retrieved from http://dataspace.princeton.edu/jspui/handle/88435/dsp014b29b837r

Latour, Bruno (2009) Tarde's idea of quantification. In Matei Candea (ed.), *The social after Gabriel Tarde: debates and assessments* (pp.187-202). London: Routledge. Retrieved from http://www.bruno-latour.fr/sites/default/files/116-CANDEA-TARDE-FR.pdf

Los Angeles Times (2016), 30 December. Retrieved from http://www.latimes.com/books/jacketcopy/la-ca-jc-cathy-oneil-20161229-story.html

MacNeil, Heather (2007). Archival Theory and Practice: between two paradigms. *Archives & Social Studies: A journal of Interdisciplinary Research, 1*, 517-545.

Manoff, Marlene (2004). Theories of the Archive from Across the Disciplines. *Libraries and the Academy, 4*(1), 9-25.

Mayer-Schönberger, Viktor, & Kenneth Cukier (2013). *Big Data: a revolution that will transform how we live, work and think*. Boston / New York: Houghton Mifflin Harcourt.

Mason, Scott (2016). *Nature of Knowledge*. Delhi: The centre for Internet & Society. Retrieved from http://cis-india.org/internet-governance/blog/nature-of-knowledge#_ftnref19.

McQuillan, Dan (2016). Algorithmic paranoia and the convivial alternative. *Big Data & Society, 3*(2), 1-12.

Morozov, Evgeny (2014). *To save everything, click here. Technology, solutionism and the urge to fix problems that don't exist*. London: Penguin.

Mul, Jos de (2002). *Cyberspace Odyssee*. Kampen: Klement.

O'Neil, Cathy (2016). *Weapons of Math Destruction. How Big Data increases inequality and threatens democracy*. New York: Crown.

Packer, Jeremy (2010). What is an Archive?: An Apparatus Model for Communications and Media History. *The Communication Review, 13*, 88-104.

Rodrigues, Nuno (2016). Algorithmic Governmentality, Smart Cities and Spatial Justice. *Justice Spatiale | Spatial Justice, 10*(July). Retrieved from http://www.jssj.org

Rose, Karen, Scott Eldridge, & Lyman Chapin (2015). *The Internet of Things: an overview. Understanding the issues and challenges of a more connected world*. Geneva (SW) / Reston, VIC: The Internet Society. Retrieved from https://www.internetsociety.org/sites/default/files/ISOC-IoT-Overview-20151014_0.pdf

Røssaak, Eivind (2016). Memory and Media. Archival tasks in the age of algorithms. In *Schetsboek Digitale Onderzoeksomgeving en Dienstverlening. Van vraag naar experiment*. Den Haag: Stichting Archief Publicaties.

Rouvroy, Antoinette, & Thomas Berns (2013). Gouvernementalité algorithmique et perspectives d'émancipation. *Réseaux, 1*(177), 163-196.

Scott, James C. (1998). *Seeing like a State. How certain schemes to improve the human condition have failed.* New Haven / London: Yale University Press.

Thomassen, Theo (1999). Een korte introductie in de archivistiek. In P.J. Horsman, F.C.J. Ketelaar and T.H.P.M. Thomassen (eds.), *Paradigma. Naar een nieuw paradigm in de archivistiek* (pp. 11-20). Den Haag: Stichting Archief Publicaties). (Translated in English (2000): A first Introduction in Archival Science. Archival Science, 1, 373-385.)

Tokmetzis, Dimitri (2012). *De Digitale Schaduw. Hoe het verlies van privacy en de opkomst van digitale profielen uw leven beïnvloeden*. Houten / Antwerpen: Spectrum.

Upward, Frank, Barbara Reed, Gillian Oliver, & Joanne Evans (2013). Recordkeeping informatics: re-figuring a discipline in crisis with a single-minded approach. *Records Management Journal, 23*(1), 27-50.

Villasenor, John (2011). *Recording Everything: Digital Storage as an Enabler of Authoritarian Governments*. (Center for Technology Innovation (Brookings)). Retrieved from https://www.brookings.edu/research/recording-everything-digital-storage-as-an-enabler-of-authoritarian-governments/

Wetenschappelijke Raad voor het Regeringsbeleid [Scientific Council for Government Policy] (2011). *iGovernment*. Amsterdam: Amsterdam University Press.

Wetenschappelijke Raad voor het Regeringsbeleid [Scientific Council for Government Policy] (2016). *Big Data in een vrije en veilige samenleving*. Den Haag: WRR. Retrieved from https://www.wrr.nl/publicaties/rapporten/2016/04/28/big-data-in-een-vrije-en-veilige-samenleving

Yeo, Geoffrey (2007). Concepts of Record (1): evidence, information, and persistent representations. *The American Archivist, 70*, 315-345.

ANNE J. GILLILAND

'The Wink that's Worth a Thousand Words': A Contemplation on the Nature of Metadata and Metadata Practices in the Archival World

Overview

As with the term 'information', 'metadata' is both ubiquitous and applied in so many ways in this digital age that without conceptual analysis and close operational definition, while it may be intuitively understood, it is essentially expressively useless. This paper addresses questions about what is to be gained philosophically and practically from a discursive examination of metadata by archival science and other recordkeeping fields, which play crucial cultural, memory, evidentiary and information roles in society. It argues that philosophically and phenomenologically such an examination is important because, visibly or invisibly, metadata is a factor that is at work in all systems and services that support such roles, and is also embedded in and envelopes every type of informational, evidentiary and cultural resource with which these fields engage. After a preliminary discussion about the definition of metadata, this paper briefly reviews the history of metadata in archival science and recordkeeping more broadly. From there it contemplates, with illustrations, the concept of metadata in terms of its various and expanding conceptualizations and instantiations, as well as some ethical, political and emerging concerns.

Introduction

British producer, musician and artist Brian Eno, participating in a panel discussion at the *Time & Bits: Managing Digital Continuity* conference organized by the Getty in 1998, talked of "the wink that's worth a thousand words ... the wink at the right moment, which everybody knows what it means, but it's much too complicated to explain" (McLean & Davis, p. 51). Eno was metaphorically alluding to the under-explicated, yet widely used term 'metadata', and his comment continues to provoke fundamental questions about what is to be gained either philosophically or practically from a discursive examination of metadata. Scholars in library and information science (LIS) have engaged in extensive philosophical treatments of the nature of information–Buckland's (1991) discussion of information-as-process, information-as-knowledge and information-as-thing, and Furner's (2014)

consideration of multiple possible approaches to answering a question such as 'what is information?' immediately come to mind–but metadata has not received the same treatment in any information field. As with the term 'information', 'metadata' is both ubiquitous and applied in so many ways in this digital age that without conceptual analysis and close operational definition, while it may indeed be intuitively understood it is essentially expressively useless, hence Eno's 'wink'. This paper addresses what is to be gained from such a discussion from the perspective of archival science and other recordkeeping fields, which play crucial cultural, memory, evidentiary and information roles in society. It argues that philosophically and phenomenologically such an examination is important because, visibly or invisibly, metadata is a factor that is at work in all systems and services that support such roles, and is also embedded in and envelopes every type of informational, evidentiary and cultural resource with which these fields engage. It briefly reviews the history of metadata in archival science and recordkeeping more broadly. From there it contemplates, with illustrations, the concept of metadata in terms of its various and expanding conceptualizations and instantiations, as well as some ethical, political and emerging concerns.

Metadata is not a New Phenomenon

Metadata is a relative neologism with respect to the inherited canon of archival principles and ideas. That canon was largely formulated in the nineteenth and first half of the twentieth century and, while increasingly challenged philosophically and practically, its strictures about archival arrangement and description continue to dominate professional thought and practice (Gilliland-Swetland, 2000). Metadata, however, is a broader concept than simply arrangement and/or description, even though the term may sometimes be used synonymously by those engaged in description and other information organization activities. In the archival field it came into currency in connection with the management of born-digital or 'electronic' records, but it would be a mistake to consider it to be something new that emerged in the digital era. In fact, metadata in various complex and evolving instantiations have always been inherent to the nature of records as they have been to any other kind of information or cultural object. And as already mentioned, the exploitation and application of metadata are also fundamental to the entire body of practices of archival science and recordkeeping more broadly, including the design of systems that create records and the structure or form of those records, as well as their appraisal and preservation (Bearman, 1989, p. 37).

The Enlightenment is often looked to as the impetus behind the development of systematic knowledge organization schemes in different disciplines. However archaeological and historical evidence indicate that keepers of records and prototypical archives developed and implemented bureaucratic organizational schemes and formulaic documentary structures in even the earliest days of written recordkeeping in the Fertile Crescent (Gilliland, 2011). Roman systems of registering and abstracting documents were precursors of registries implemented across subsequent empires of colonial powers, mercantile enterprises, and the reaches of the Roman Catholic and Orthodox churches. Registries were integral to organizational workflow and control, and ensured the arrangement and description of bureaucratic and religious records through registration and classification, identified versions of records, and tracked their subsequent circulation. Latterly

instantiations of registries are still widely in use around the world in physical and digital forms and have inspired many of the requirements and approaches that are embedded in the current ISO recordkeeping metadata standards.

Jacob von Rammingen, of the German tradition, wrote in 1570 what is considered to be one of the first treatises in archival science or *Wissenschaft*. He asserted that no two registries would likely be the same and underscored the complexities of the entities, tools and activities associated with archival organization:

> "In our registry ... we need the following genera. First, one for what is received, and one for what is emitted. And also for notes and announcements. Then for copies, originals, drafts and abstracts. Then for comments and supplements. Finally for indexes and repertories, and also for securities and responses. As for other books and registers ... the zealous and industrious analyzer and segregator will, in their eases, quickly realize what books and registers he needs for this." (p. 97)

However, with concern for his own job security, teaching income and potentially personal safety if the records in his charge were to become too easy to identify or locate, he refused to divulge much more detail in the treatise, stating that "Anyone wishing more information about this can pay the tuition fee, and then learn it by seeing with his own eyes, and hear the oral teaching of these secret 'kabbalistic' traditions." (p. 98). Head, in his study of early modern archives in Switzerland, has found a "seeming progression from listing to mapping to taxonomy" (2016, p. 433) in organizational approaches. His research suggests, in line with Rammingen's observation, that a heterogeneity of descriptive approaches were in play during this period, and he cautions against over-generalizing their nature. Other historians who have examined descriptive systems during and after the Reformation also note divergences and overt politics of organizational schemes connected to how archives were being used to support the Protestant and Catholic movements, institutions and theologies of the time (Head, 2010). For example, cosmographers in the sixteenth and seventeenth century Spanish court were charged with devising data collection forms and cosmographic schemes that could incorporate the new knowledge and natural history being encountered as Spain built its empire in the so-called Indies into existing Catholic cosmography. These structures and ontologies became the metadata that governed the nature and the interpretation of the records that would be perused by the king and eventually gathered in the extensive Archives of the Indies (Portuondo, 2009).

Contemporary Constructions of Archival and Recordkeeping Metadata

When the term 'metadata' was first used, by the geospatial, data management and systems design communities, it referred to the internal and external documentation necessary for the identification, representation, interoperability, technical management, performance and use of data contained in information or other automated systems (Gilliland, 2016, p. 1). By the 1990s the term was widely adopted by professionals engaged in the organization of information and especially in bibliographic description to refer to catalog records and other forms of value-added resource description that they were creating. More than a case of 'old wine in new bottles,' it provided a broader and more interdisciplinary way to conceptualize their work and the new standards and other descriptive tools they were developing in the digital and networked era. Since the development of international standards

for archival description, the archival profession has interacted increasingly closely with the worlds of both bibliographic and museum description.

Taking both of these trajectories into account, multiple ways have been proposed, therefore, to identify and operationalize metadata in information and cultural heritage contexts (e.g., administrative, descriptive, preservation, technical, use) and to characterize it (e.g., source of metadata, method of metadata creation, nature of metadata, metadata status, structure, semantics and level) (Gilliland, 2016). Archival and recordkeeping preoccupations today engage explicitly or implicitly with all of these ways. Distinctively, archivists and other recordkeepers are concerned with bureaucratic accountability and transparency, as well as with the preservability of legal, historical and cultural evidence. These concerns set a particularly high bar for the continual management of trustworthy metadata necessary to audit recordkeeping systems and practices and validate and (re)produce records. Nevertheless, despite archival science being a field that is given to reflecting upon and developing its theoretical base, the extent to which archives and recordkeeping were overtly engaged with metadata beyond description was not appreciated until the archival science and other recordkeeping fields had to confront the management of electronic records. When records were predominantly in paper form, their manifestations and nature seemed to be more self-apparent and were less subject to conceptual analysis about their identity and constitution. One important exception to this assertion should be noted, however, and that is the diplomatic analysis of the genesis, form, transmission and documentary context of individual documents. The development and application of diplomatic techniques, notably in and after the seventeenth century, initially sought to determine the authenticity of mediaeval charters and thus the validity of legal claims contained therein. In the nineteenth century diplomatic techniques became more specialized as they expanded to support the historical analysis and authentication of many other common and emerging types of records though an examination of elements such as the acts, actors, form, dates, copies and versions, and seals associated with the document in hand in addition to the likely veracity of the information that it contained. Diplomatic ideas and techniques were extended to twentieth century documents (Carucci, 1987) and by the end of the twentieth century by Duranti in the form of 'contemporary archival diplomatics' to address aggregations of records (i.e., rather than individual documents or instances of record types) as well as records that had been born-digital (1998).

With electronic records a key consideration is that there is not necessarily any physical object in hand to manage, describe or make available, and sometimes there is only the capacity to render or recreate a record virtually:

> "[Electronic records] are [often] heterogeneous distributed objects comprising selected data elements that are pulled together by activity-related metadata such as audit trails, reports, and views through a process prescribed by the business function for a purpose that is juridically required." (Gilliland-Swetland & Eppard, 2000)

In other words, they are intellectually complex and contingent objects to identify and move forward through time and migrations without compromising their authenticity, and thus need to be described as a conceptual as well as a virtual object, and in relation to all of their contingencies. The first usages and glossary definitions of 'metadata' in the field unsurprisingly therefore did not derive from how metadata was being conceived in the information organization fields. Rather they emanated

out of understandings that had developed from analyzing and (re)designing government electronic recordkeeping systems in the 1980s so that they could capture and exploit both static and dynamic process metadata necessary for evidentiary, accountability and preservation purposes (United Nations ACCIS, 1990). In 1993 Wallace argued that a metadata systems approach could provide solutions to many of the problems that had been identified with managing records produced by electronic systems. He synthesized many of the advantages of this approach that had been recognized by those engaged in electronic records management:

> "(1) capture and preservation of records context (evidence);
> (2) preservation of systems and record structure;
> (3) generation and retention of relevant descriptive information;
> (4) incorporation of appraisal and disposition data;
> (5) life cycle management of records;
> (6) preservation and migration of system functionality; and
> (7) creation of inventory/locator systems for organizational information resources" (p. 88).

Wallace subsequently also noted the definition used by the 1989 Society of American Archivists (SAA) Working Group on Standards for Archival Description:

> "[the] process of capturing, collating, analyzing, and organizing and information that serves to identify, manage, locate, and interpret the holdings of archival institutions and explain the contexts and records systems from which those holdings were selected" (Wallace, 1996, p. 17-18).

He argued for the potential for the automated creation and capture of descriptive metadata out of appropriately designed electronic recordkeeping systems, thus beginning the embedding of a metadata consciousness across all areas of archival and recordkeeping activity in order to support "record identification, access, understandability, interpretation, authenticity, and ongoing management" (Wallace, 1993, p. 100; Wallace, 1996, p. 18; Hedstrom, 1993).

The UBC and InterPARES research projects (Duranti, 1997; Interpares.org; Interparestrust.org) have applied Duranti's diplomatics approach to delineate mechanisms for ensuring the reliability and authenticity of electronic records, focusing on intrinsic and extrinsic elements of documentary form, annotations, context (encompassing juridical-administrative, provenancial, administrative, procedural, documentary and technological) and medium (MacNeil, 2016). InterPARES research found that many of the requirements diplomatically established for creating reliable and preserving authentic electronic records:

> "... could potentially be implemented through metadata and archival description, particularly such aspects as identity, linkages, documentation of documentary forms, juridical requirements, business rules and technical procedures, access privileges, establishment of the authoritative record when multiple copies exist and transfer of relevant documentation." (Duranti & Preston, 2008, p. 13)

Testing this assertion, InterPARES developed a metadata specification model for its Chain of Preservation (i.e., the records life cycle) model. They defined 'metadata' as *a machine or human-readable assertion about a resource relating to records and their resources*, and descriptive metadata was defined as those categories of metadata carried forward to be used as evidence for archival description. Speaking to the

ubiquity, heterogeneity and multifunctionality of metadata, this test identified 137 different metadata assertions (i.e., different instances of types of metadata), and 16 types of assertions. Two types cut across all stages of the lifecycle, one cut across two stages, and the other fifteen were evidenced only in one stage. (Gilliland et al., 2008, p. 31).

In 1999, the Australian Recordkeeping Metadata Schema (RKMS) concisely identified the objects that are the primary focus of archivists and other recordkeepers as a set of entities such as records, recordkeeping, agents, business process and mandates *as well as* the various relationships that might exist between each at any point in timespace (McKemmish et al.). The developers of the schema assert that these are the objects that need to be described at relevant points in their lives, from the moments of systems design and records creation onwards. The task of that description is to ensure that these objects can be understood semantically and epistemologically, and trusted bureaucratically, juridically and socially. Additionally they must help the user to assess the object with regard to its authoritativeness, authenticity and reliability; they must place it within its broader documentary context or archival bond (e.g., within a given fond); identify it (and here it would be interesting to look at the degree of consonance with ideas developed in the LIS field of cognitive authority as well as of 'relevance' and 'aboutness' (Wilson, 1983; 1968)); and support potentially infinite interpretations and understandings of what the objects being described reveal about dynamic constructions of identity, memory, and truth.

All of these abilities are, of course, in turn contingent upon the degree, nature and ultimately the trustworthiness of the associated metadata, so much so that the management, elimination and preservation of digital metadata through some form of trust regime has become a growing professional concern (Gilliland, Rouche, Evans and Lindberg, 2005). If trusting a record means the existence of trustworthy metadata, how then does one manage constantly accruing metadata in such a way that it maintains its trustworthiness? This may prove to be one of the most pressing questions facing these fields in the face of escalating amounts of metadata in the digital world and also growing political distrust in records and data. One suggested approach is to develop a metadata management scheme whereby redundant metadata would be eliminated and notarized summaries would be prepared for large aggregations of metadata, which could then also be discarded. Another possible approach would be more ecological–relying upon organizations to invest more in sustaining the trustworthiness of those records and other objects that they wish or need to trust and preserve. However both approaches raise important ethical issues related to equity and the resource capacity of different institutions and communities to manage relevant necessary metadata over the long-term that meets professionally and legally acceptable practices for preserving its authenticity.

One additional observation should be made at this point. As Wallace predicted, metadata has proven to be essential in recordkeeping, and to have more and more potential as new digital capabilities emerge, but it is not everything. Recent developments in digital forensics rely not on metadata but instead on very physical properties of the medium on which records are inscribed. Nevertheless, interpreting the trace inscriptions detected through digital forensics also must draw upon existing trusted contextual metadata.

Finally, although the concept of metadata in the archival science and other recordkeeping fields has been defined primarily in the context of electronic records – and theoretical developments regarding born-digital materials have largely been focused on records created by bureaucratic activities and systems – it is important to bear in mind, as stated at the outset of this paper, that metadata is inherent to all information and cultural objects. Electronic records research and development have helped us to understand the roles and manifestations of metadata, but everything we have learned can be applied to records and other archival materials regardless of their genesis or form. For example, Yeo makes the point that organic accumulations and artificial collections may not be mutually exclusive aggregations, even though theory and practice tends to draw quite a rigid distinction between the two (Yeo, 2010; 2012a; 2012b). Artificial collections, as well as aggregations of personal papers are also the products of human activity and also have inherent metadata that speak to their genesis, nature, transmission, aggregation and so forth. Moreover, as we move forward and use value-added linking metadata such as RDF, as well as automatically exploiting and compiling inherent metadata elements in archival, digital humanities and big data initiatives, we will be moving into a new universe of meta-fonds and meta-collections and meta-metadata.

Characteristics of Archival and Recordkeeping Metadata

Notwithstanding Eno's admonition that metadata is too complicated to explain, at this point we can observe that at its most straightforward, archival and recordkeeping metadata can be thought of as encompassing anything and everything that relates to an object of interest within these fields (i.e., any entity or relationship) that is not the object itself. This metadata accumulates and envelopes the object as it moves through time. Metadata also operates in various one-to-one and one-to-many relationships. For example, metadata can relate to an individual object, or to an aggregation of objects. Structural, epistemological and ontological relationships exist between different types and instances of metadata, in addition to the relationships between the kinds of entities identified in RKMS. For example, there are multiple organic and cumulative relationships between the provenancial, descriptive, preservation, interpretative, affective, performative, and 'making' and 're-making' metadata layers that accumulate around and between the original record 'objects' as well as their versions and copies and the fonds within which they are filed. Originals, copies and parts of a record can also participate in new archival formations, with various familial-type relationships (parent-child, sibling, spouse, cousin, generational) that potentially have certain resonances with the notion of bibliographic families first articulated by Wilson (1986). Digital capabilities only expand the ways and moments in and at which this can occur.

We might further consider dividing metadata into two types. The first we might call 'trace metadata', that is metadata that is indigenous to the records and the system and agents that created and handled them. Trace metadata is often created simultaneously with the actions that it reflects, and accumulates across actions, for which it serves as evidence. That evidence might be called upon in legal and audit processes, for institutional decision-making and self-knowledge, and for historical analyses, among other uses.

We might call the second type 'value-added metadata', that is, description and arrangement schemes imposed by an archivist, other records expert, or user, or automatically captured and repurposed trace metadata. Such metadata describes the object in hand, augments what is not self-evident of the circumstances of its creation and handling with the intent of making apparent not only its evidential aspects but also the information it contains "about particular persons, situations, events, conditions, problems, materials, and properties in relation to which the question of action comes up" (Schellenberg, 1956). It makes the object more usable for subsequent users, who may bring different backgrounds from those of the creators of the object, and who may wish to use the objects for completely different purposes than those for which they were designed. For example, a handwritten ledger kept by a clerk of courts listing applications of Intent to Naturalize would likely have an individual entry corresponding to each application and listing the applicant's name, city or county of current residence, date of birth, date and port of arrival in the country and a code linking the entry to the Intent to Naturalize form. The primary access point to the ledger would be the date when the application was made and the ledger would list the applications chronologically in the order received. A genealogist trying to find an entry for an ancestor in these ledgers (and thereby, to retrieve the relevant application form) would need to know the date upon which the application would have been filed – an unlikely data point for the genealogist to know. Value-added description, therefore, might consist of abstracting the ledger to create a database index to the ledger that would permit random searching on any of the data elements contained in the ledger entries as well as the ability to compile various statistical reports on application rates and applicant demographics.

With a continuum approach, which "maps the creation of records as traces of actions, events and participants, their capture into systems [broadly defined]", their organization into an archive and their pluralization beyond the confines of their original creator and the archive (McKemmish, 2016, p. 139), both of these types of metadata are created at all points in the life of the object. With the life cycle approach, value-added archival metadata would usually be created post-archival accession. Beyond the obvious temporal differences in when the metadata are created in these two approaches, there is an important qualitative difference between value-added data created close to the time when the action with which it is associated is performed (e.g., creation, active bureaucratic use or migration) and that created later, sometimes many years later, without firsthand knowledge of those actions but with the benefit of hindsight. Although the latter value-added metadata are still created with reference to the object of description itself, they are also based upon the archivist's assessments of the past and with knowledge of intervening events that might cast the resources in a somewhat different light from that in which they were perceived originally or at an earlier date. Again, it would be interesting to contemplate this phenomenon with reference to that of second-hand knowledge as described by Wilson (1986) with reference to bibliographic description. It should also be noted that in legal theories of evidence, the closer in time to an action that a record is created, the more likely it is to be reliable. By implication, this also would be the case for those metadata that are essential to supporting the reliability of that record. Ultimately, even in a life cycle approach, however, these types are non-exclusive because, depending upon one's standpoint, all metadata could be read as 'metadata as trace' in the sense that it is always

evidence of what has happened to the object over time and of how it has been interpreted and valued. This discussion also neglects a third, less tangible category of *use and interpretation* metadata–that which is brought to bear by a user (e.g., scholar, genealogist, lawyer, artist) in locating and using or exploiting an object in the form of pre-existing knowledge, purpose, and intellectual, ethical, political or artistic stance.

These kinds of categorizations are somewhat crude ways of conceptualizing metadata that better support practical than philosophical understanding. From a more conceptual perspective, can we identify any other characteristics that might help us to apprehend the nature and role of metadata? The following discussion suggests a few areas for further contemplation:

Metadata is what makes something a record: Fundamentally, because they are co-constructions, any discussion of the nature of metadata requires an understanding of the nature of a record as a conceptual and juridical as well as an information object, and vice versa. The circular definition of metadata as 'data about data' that was often cited in other information fields has proven to be an overly reductivist way to talk about the kinds of highly contingent objects with which the body of archival and recordkeeping theory and practice grapples. A record can convey information, but its primary distinguishing characteristic is its evidential capacity to bear witness, formally or informally. How well it does that depends on how much metadata is associated with it, as well as how trustworthy that metadata is. Without metadata associated with its various contexts, there may be a poor record, or even no record at all, merely discrete pieces and points of data. In other words, a record must have metadata, and the sufficiency and reliability of that metadata speak to the quality of the record in terms of its comprehensiveness and trustworthiness.

Metadata can be a record: Conceptually, metadata is a paradox, for besides serving as a running commentary on, and validator and representation of the record(s), it can also be construed as a record in its own right (at various levels of completeness and granularity) and will have its own associated metadata. By extension, metadata as records need their own metadata not only to be reliable, but also to be understandable and referential. For example, embedded in the wall of the entrance to one of the Oxford colleges is a plain concrete slab which reads as follows "This Foundation Stone was Laid by Her Majesty the Queen 2 May 1968). In this case, explanatory metadata disambiguating to which Queen it refers, and linking the placing of the foundation stone to the broader history of the college, has been presumed to be unnecessary since it will instead be contributed by the contextual knowledge brought by the reader of the stone. As already discussed, explications of metadata in archives and recordkeeping often fail to take into account the metadata that users bring or need to bring to objects in their sensemaking process, and such metadata may not always be relied upon to exist, or to travel well across time and knowledge bases. For example, a visitor from another country and in a future time may not easily be able to figure out that it was British Queen Elizabeth II who laid the foundation stone. Metadata for films provide us with a rather different example of meaningfulness and understandings–viewers are expected to understand how the ordering of actor credits can indicate such things as the standing and pay of actors. Uncredited cameos in cast lists may suggest a lack of confidence in the quality of the film or their own performance, on the part of the unlisted actor. *Metadata can and must be used to demonstrate the fixity of records even while the*

dynamic accrual of metadata relating to the record is continual. As with the archive, metadata has no end, and its beginnings may actually precede the records to which it relates (e.g., metadata embedded in systems design). Even if physically fixed in some tangible manifestation, the record has never been a static intellectual object and archiving as an activity cannot bind time, it can only capture and annotate recordkeeping and record moments. As Eno states with regard to how a record is pluralized, it is different each time an action is performed on or with, and every action performed on or with a record adds to its metadata:

> "… what we're talking about is not evolving the original, we're talking about keeping the original and adding layers of annotation, you might say, to layers of commentary to it. So it's slightly different from a generative process where you plant a seed and it turns into something else. We want the seed to stay intact as well." (p.57)

So, while the fixity of the archived record is a condition for demonstrating its authenticity, and that fixity is demonstrated through metadata such as audit trails and documentation of migrations, it remains conceptually impossible to fix either a record or its metadata. McKemmish explains how continuum-based recordkeeping addresses this additional paradox:

> "Pluralisation involves disembedding the record from its multiple organisational and/or personal contexts and carrying it through spacetime. Thus recordkeeping processes fix the content and structure of records, preserve their integrity by ensuring they are tamper-proof, and link them to ever-widening layers of rich metadata about their multiple contexts of creation and use. This enables them to be accessed, used and interpreted in other spacetimes. In continuum terms, while a records content and structure can be seen as fixed, in terms of its contextualisation, a record is 'always in a process of becoming' (McKemmish, 2016, p. 139).

This in turn points to the potential to use metadata to facilitate rupturing the strictures of structure so that people, individually and collectively, can engage with archives and their holdings in entirely new ways.

Metadata together with and apart from the objects with which they are associated can have social lives: Chinese artist and activist Ai Weiwei's controversial 1995 piece "Dropping a Han Dynasty Urn", where he deliberately dropped an urn that was deemed to be priceless, sought to raise questions about the life and preservation of a cultural heritage object, the assumption that is should remain fixed and stable, and when, if ever, it might transform into another object. Influenced by the work of scholars such as Kopytoff and Appadurai (1983), archival scholars such as Caswell (2014) and Carbone (2017) have begun to explore the social lives, cultural biographies and movements of archives and recordkeeping objects and their metadata in order to explore these questions of fixity and stability, and to expose issues of agency and affect in the generation of new and derivative objects out of archival objects. Caswell traced the subsequent trajectories of 'mug shots'–photographs taken of prisoners of the Khmer Rouge before they were executed and their copies--across collections, exhibitions, publications and digitization projects (Caswell, 2014). Carbone conducted an ethnographic investigation of how artists in residence in Portland City Archives were inspired by one particular collection and moved and transformed objects they found there into new spaces, interpretations and ultimately, new objects (Carbone, 2017). In each case, some parts of the pre-

existing metadata associated with the objects travelled with them and were modified
and augmented. Metadata can also be separated from their associated object and
transformed over time. For example, metadata for an object can be incorporated
into a catalog for an exhibition that in turn may be reformulated into a monograph
on the same subject.

New views through metadata: Metadata have the potential to reveal the extent and
limits of different archival universes by providing the infrastructure that facilitates
meta-compilation, visualization and analysis. Diasporas of records can be discerned
and linked, and their contents can be indexed, abstracted and compiled to provide
new ways to understand and imagine the past. Similarly, they can support the
presentation of meta-level views of the scope, functions and infrastructure of
recordkeeping.

Metadata can indicate meaningful absences of records and ruptures in recordkeeping:
Archives operate in an often-contested world where the complete set of original
records rarely exists—maybe it was never created, maybe it or parts were not deemed
worthy of keeping, were defensively destroyed, or were hidden or lost. Metadata
helps us to explain and draw inferences in the face of such absence. It not only
supports objects that are present or in existence, it can also explain and interpolate
what is missing. There are many historical examples of records and the contents of
archives being lost or destroyed but their metadata, in forms such as classification
schemes, catalogs and abstracts survive (MacNeil, 2017). From that surviving
metadata we can infer something of what is missing as well as what its intellectual
and physical arrangement might have been, and even perhaps the contemporary
value or power attributed to it and what might have motivated its destruction. For
example, German record classification schemes were introduced in the Nuremberg
trials to stand for records destroyed by German authorities at the end of World War
II; and metadata for photographs of demolished Bedouin villages in Gaza have
become important for claims actions since they testify to their location when places
have been erased or renamed. Foscarini argues that records exist in a rhetorical
landscape and this allows us to see workarounds or irregularity in recordkeeping.
Beyond diplomatics, she recommends the use of genre theory to analyze the contexts
of records, contexts that are captured by means of metadata (2012). Metadata, for
example that relating to work flow and juridical requirements, can similarly be used
to indicate where a record has not been created but should have been, or should be.
Moreover, metadata can be the traces around which not only previously existing, but
also imagined records can be hung (Gilliland & Caswell, 2017). Empty metadata
fields may also be meaningful. For example, an Armenian curator talks of
attempting to use retrospectively Dublin Core to describe museum objects dispersed
during the Armenian genocide and being faced with fields that will never be filled
because they cannot be (Hovhanissyan, lecture, 2017).

The Politics and Ethics of Metadata in a Plural World

Eno's 'wink' was not only to the phenomenon of complex and cumulative
instantiations and versions of metadata in the digital world, but also to how
contemplations of metadata and metadata characteristics surface professional
belief and value systems inherent in the long-term and efficient management of
objects under their purview. In the case of archives and recordkeeping, these might
include, for example, tensions between investing in metadata to support

bureaucratic requirements (e.g., ensuring accountability, transparency and effectiveness) or cultural, scholarly or even human and civil rights imperatives when resources are constrained. While ethical considerations regarding metadata in the twentieth century might also have dwelled on how to document and control levels of access, meet privacy requirements, redact content and implement other privacy and security controls, more recently professional debate has centred around issues of equity, power, voice and vulnerability, raising questions such as what is authoritative metadata? Who gets to create metadata? Whose interests are represented in the design of metadata standards and systems that support metadata creation? Whose words are used in metadata and who has the power to name? Whose presence in the records should be emphasized and who should speak for that presence? How much metadata should be created and at what levels of granularity (since granularity often determines utility for particular uses)? For whom and for what purposes should value-added metadata be created? And can or should metadata be proprietary? Another important area of concern focuses on disinformation and the use of metadata deliberately to mislead or to promote particular political narrative, especially in the digital environment.

A number of more applied ethical questions lie at the heart of current metadata practices in the archival and recordkeeping fields and challenge how they will move toward more accessible, equitable and networked digital and global futures. For example, if, in its simplest construction, metadata can be considered to provide essential context for records and their use, where does this context begin and end (if at all) and how do professionals manage it to ensure that it does not eventually overwhelm the record or records to which it pertains? How might expanded conceptualisations of key metadata elements such as provenance play in pluralizing understandings of and acknowledging rights in records? What rights do those who are implicated in the record have as to what metadata is captured, assigned, compiled and publicly disseminated about them? And what types of metadata and metadata-based tools should be made available to the public to assist in metadata compilation and manipulation within and across archival holdings?

It is beyond the scope of this paper to attempt to try to answer these questions, but it is interesting to note that we can see ethical concerns surfacing, even as early as von Rammingen's treatise. In what he views to be a decadent era, he is clear about the ethics underlying the work of organizing the registry--that it can hold people accountable by exposing or detecting incorrect or falsified information and records; and that it can ensure that those who have lost, because of wars or suffering, relevant charters, can still recover the textual content of those documents through the registry, thus evidencing their rights (pp. 7-8).

Final Thoughts

Metadata are endless, as are archives. Their extent is only limited by humanity's resources, technical capabilities, imagination, and will. No universal hard and fast distinction between record and metadata has ever been possible to make--and it is not so much that this is different from the cases of other information objects and professions so much as it is something of which archivists and other recordkeepers are deeply conscious. Non-digital practice and other information fields can all benefit from applying a close analysis of metadata. Given the centrality of metadata to archival science and recordkeeping theory and practice, however, and its

particular relevance to the central concepts of 'context' and the creation and preservation and (re)production of 'evidence' across time, it is essential that these fields engage in such a contemplation and in doing so, have the potential to contribute provocative and far-reaching philosophical work on the subject to the broader informational, cultural, juridical and bureaucratic realms within which they are situated.

Glossary

Recordkeeping: Encompasses all aspects of the creation, management and use of records and their associated metadata across space, time, agents, mandates, motivations and manifestations. It thus subsumes those aspects traditionally considered to be the professional field of archival science or archivistics as well as records management. Since this is not always understood as such outside continuum contexts this essay uses the construction (really a misconstruction) 'archival science and recordkeeping' to underscore that it is taking a broad view on metadata phenomena.

Records continuum: "Encompasses a range of intertwined recordkeeping and archival processes and activities carried out by records managers and archivists for current, regulatory, and historical recordkeeping purposes. These purposes include the roles that recordkeeping plays in and through space and time in governance and accountability, remembering and forgetting, shaping identity and providing value-added sources of information. In classificatory terms 'recordkeeping' in this usage subsumes records management and archival administration. It also encompasses the personal and corporate recordkeeping activities undertaken by individuals in their everyday lives, in families, work or community groups, and in organisations of all kinds" (McKemmish, Upward & Reed, 2009, 4448)

Literature

Appadurai, A. (Ed.) (1986). *The social life of things: Commodities in cultural perspective*. New York: Cambridge University Press.

Bearman, D. A. (1989). *Archival methods*. Pittsburgh: Archives and Museum Informatics.

Buckland, M. K. (1991). Information as thing. *Journal of the American Society of Information Science, 42*(1), 351-360.

Bunn, J. A. (2016). Mind the gap: Quality from quantity. In *Proceedings of the 2016 IEEE International Conference on Big Data, BigData 2016, Washington DC, USA, December 5-8, 2016* (pp. 3240-2344). https://doi.org/10.1109/BigData.2016.7840980

Carbone, K. (2017). *Moving records: Artistic interventions and activisms in the archives*. Ph.D. dissertation at the University of California, Los Angeles. Retrieved from http://escholarship.org/uc/item/1bt983g5

Carucci, P. (1987). Il documento contemporaneo. *Diplomatica e criteri di edizione*. Rome: Carocci.

Caswell, M. (2014). *Archiving the unspeakable: Silence, memory, and the photographic record in Cambodia*. Madison, WI: University of Wisconsin Press.

Duranti, L. (1998). *Diplomatics: New uses for an old science*. Lanham, MD: Scarecrow Press and Society of American Archivists.

Duranti, L. (1996). The protection of the integrity of electronic records: An overview of the UBC-MAS Research Project. *Archivaria, 42*, 46-67.

Foscarini, F. (2012). Diplomatics and genre theory as complementary approaches. *Archival Science, 12*.

Furner, J. (2004). Conceptual analysis: A method for understanding information as evidence, and evidence as information. *Archival Science, 4*.

Furner, J. (2014). Information without information studies. In F. Ibekwe-SanJuan & T. M. Dousa (Eds.), *Theories of information, communication, and knowledge. Studies in history and philosophy of science 34*. DOI 10.1007/978-94-007-6973-17.

Gilliland, A. J. (2011). Reflections on the value of metadata archaeology for recordkeeping in a global, digital world. *Journal of the Society of Archivists, 32*(1), 97-112.

Gilliland, A. J. (2016). Setting the stage. In M. Baca (Ed.), *Introduction to metadata: Pathways to digital information*, (3rd ed.) (pp. 1-20). Los Angeles: Getty Publications. Retrieved from http://www.getty.edu/publications/intrometadata/setting-the-stage/

Gilliland, A. J., & Caswell, M. (2016). Impossible archival imaginaries and imagined records. *Archival Science, 16*, 53-75. DOI: 10.1007/s10502-015-9259-z.

Gilliland, A. J., Rouche, N., Evans, J., & Lindberg, L. (2005). Towards a twenty-first century metadata infrastructure supporting the creation, preservation and use of trustworthy records: Developing the InterPARES2 Metadata Schema Registry. *Archival Science, 5*, 43-78.

Gilliland-Swetland, A. J. (2000). *Enduring paradigm, new opportunities: The value of the archival perspective in the digital environment*. Washington, DC: Council on Library and Information Resources.

Gilliland-Swetland, A. J., & Eppard, P. (2000). Preserving the authenticity of contingent digital objects. *DLib Magazine, 6*. Retrieved from http://www.dlib.org/dlib/july00/eppard/07eppard.html

Head, R. C. (Ed.). (2010). Archival science, special issue Interdisciplinary essays on European knowledge culture, 1400-1900. *Archival Science, 10*, 3.

Head, R. C. (2016). Historical case studies of pre-modern European archives: A comparative approach. In A. J. Gilliland, S. McKemmish & A. J Lau (Eds.). *Research in the Archival Multiverse* (pp. 433-455). Clayton, VIC: Monash University Publishing. http://dx.doi.org/10.26530/OAPEN_628143

Hedstrom, M. (1993). Descriptive practices for electronic records: Deciding what is essential and imagining what is possible. *Archivaria, 36*, 53-63.

Kopytoff, I. (1986). The cultural biography of things: commodification as process. In A. Appadurai (Ed.). *The social life of things: commodities in cultural perspective* (pp. 64-91). Cambridge (UK): Cambridge University Press.

MacLean, M. & and Davis, B. H. (Eds.). (1998). *Time & bits: Managing digital continuity*. Los Angeles: Getty Publications.

MacNeil, H. (2017). Catalogues and the collecting and ordering of knowledge (1): ca. 1550-1750. *Archivaria, 82*, 27-53.

MacNeil, H. (2016). Contemporary archival diplomatics as a method of inquiry: Lessons learned from two research projects. In A. J. Gilliland, S. McKemmish & A. J Lau (Eds.). *Research in the Archival Multiverse* (pp. 724-761). Clayton, VIC: Monash University Publishing. http://dx.doi.org/10.26530/OAPEN_628143

McKemmish, S. (1994). Are records ever actual? In S. McKemmish & M. Piggott (Eds.). *The records continuum: Ian Maclean and Australian Archives first fifty years* (pp. 187-203). Clayton, VIC: Ancora Press in association with Australian Archives.

McKemmish, S. (2016). Recordkeeping in the Continuum: An Australian tradition. In A. J. Gilliland, S. McKemmish & A. J Lau (Eds.). *Research in the Archival Multiverse* (pp. 122-160). Clayton, VIC: Monash University Publishing. http://dx.doi.org/10.26530/OAPEN_628143

McKemmish, S. (2005). Traces: Document, record, archive, archives. In S. McKemmish,
M. Piggott, B. Reed, & F. Upward (Eds.), *Archives: Recordkeeping in society* (pp. 1-20). Wagga
Wagga: Charles Sturt University Centre for Information Studies.

McKemmish, S., Acland, G., Ward, N. and Reed, B. (1999). Describing records in context in
the Continuum: The Australian Recordkeeping Metadata Schema. *Archivaria, 48*, 3-43.

McKemmish, S., Upward, F. H., and Reed, B. (2009). The Record Continuum Model. In
M. J. Bates and M. Niles Maack (Eds.), *Encyclopedia of library and Information sciences*, 3d. ed.
(pp. 4447-59). New York: Taylor and Francis.

Portuondo, M. (2009). *Secret science: Spanish cosmography and the New World*. University of
Chicago Press.

Schellenberg, T. R. (1956). Evidential values: Reasons for test of evidential values. *Bulletins of
the National Archives, 8. Washington*, DC: The National Archives. Retrieved from https://www.
archives.gov/research/alic/reference/archives-resources/appraisal-evidential-values.html

Strömberg, JBLD (Trans.). (2010). *The earliest predecessors of archival science: Jacob von
Rammingen's two manuals of registry and archival management, printed in 1571*. Lund: Wallin &
Dalholm.

United Nations Advisory Commission for the Co-ordination of Information Systems (ACCIS).
(1990). *Management of electronic records: Issues and guidelines*. New York: United Nations.

Wallace, D. A. (1995). Managing the present: Metadata as archival description. *Archivaria, 39*,
11-21.

Wallace, D. A. (1993). Metadata and the archival management of electronic records.
Archivaria, 36, 87-110.

Wilson, P. (1983). *Second-hand knowledge: An inquiry into cognitive authority*. Westport, CT:
Greenwood.

Wilson, P. (1968). *Two kinds of power: An essay on bibliographical control.* Berkeley: University of
California Publications: Librarianship, No.5.

Yeo, G. (2012a). The conceptual fonds and the physical collection. *Archivaria,73*(Spring),
43-80.

Yeo, G. (2012b). Bringing things together: Aggregate records in a digital age. *Archivaria, 74*,
43-91.

Yeo, G. (2010). *Nothing is the same as something else: Significant properties and notions of identity
and originality*. Preprint retrieved from http://discovery.ucl.ac.uk/1318039/3/1318039.pdf

GIOVANNI MICHETTI

Provenance in the Archives: The Challenge of the Digital[*]

Overview

The Principle of Provenance is a pillar of Archival Science. In its very early stages it mostly meant not to intermingle documents from different origins. This view has been challenged in the past fifty years: archival provenance has moved from a simplistic one-to-one relationship to a multi-dimensional concept based on a network of relationships between objects, agents and functions. The digital environment has posed new and unpredictable challenges: digital objects are often aggregations of several different pieces, and it is extremely easy to mix and re-use them, which makes it difficult to trace their provenance. Cloud computing has complicated the picture further. However, new technologies help us to cope with such complexity. Resource Description Framework (RDF) and ontologies can be used to represent provenance in a granular and articulated way that was not even conceivable in the past, giving us the opportunity to review and refine established practices and concepts.

Introduction

The International Council on Archives (2007) has defined provenance as:

> [t]he relationships between records and the organizations or individuals that created, accumulated and/or maintained and used them in the conduct of personal or corporate activity. Provenance is also the relationship between records and the functions which generated the need of the records. (p. 10)

In other words, archival provenance refers to the origins, custody, ownership and use of archival objects. This concept is the basis for the Principle of Provenance, a pillar of Archival Science, which prescribes that archival documents should be arranged according to their provenance in order to preserve their context, hence their meaning. This is a simplification of a complex concept that has been investigated and debated by many scholars since the nineteenth century. In its very early stages, the Principle of Provenance mostly meant not to intermingle

---

[*] This essay is going to be published in 2018 in *Authority, Provenance, Authenticity, Evidence. Proceedings of the Conference on Authority, Provenance, Authenticity, Evidence held in Zadar, Croatia, 25 to 18 October 2016. Selected papers*, edited by Ann Gilliland, Mirna Willer and Marijana Tomić (Zadar: University of Zadar, 2018). Part of the content of this essay is drawn from Giovanni Michetti, "Provenance: An Archival Perspective," in *Building Trust in Information: Perspectives on the Frontiers of Provenance*, edited by Victoria L. Lemieux (Cham, CH: Springer International Publishing, 2016), pp. 59-68.

documents from different origins: "[r]assembler les différents documents par fonds, c'est-à-dire former collection de tous les titres qui proviennent d'un corps, d'un établissement, d'une famille ou d'un individu, et disposer d'après un certain ordre les différents fonds."[1] However, maintaining the identity of a body of records as a whole is not limited to identifying its distinctness in relation to other records. Archivists soon recognized that the internal structure of such a body also shapes the identity of a fonds, and thus they established the Principle of Original Order, a corollary of the Principle of Provenance. According to this principle, records should be maintained in the same order in which they were placed by the records' creator. The underlying idea was that an archive "comes into being as the result of the activities of an administrative body or of an official, and […] it is always the reflection of the functions of that body or of that official." (Muller, Feith, Fruin, 2003, p. 19) In other words, provenance initially assumes a very physical connotation: it refers to a specific group of records, located somewhere in the repository, and arranged in a certain physical order. It is the real thing.

Fifty years ago, such a conception was challenged by Peter Scott who, in a seminal article (Scott, 1966, p. 493-504), laid the basis for a further refinement of the Principle of Provenance. He highlighted that, in general, archives are not the result of a single creator who performs a set of specific functions. They are, rather, the outcome of complex processes where different agents may act as creators. Functions change, merge and disappear; and the internal structure of the records is the result of recordkeeping activities that may have little relationship with the business activities of the creators. By extension, the structure of an archives may have little or no correspondence with the structure of the creating organization. This approach led to a new definition of the concept of provenance as it is now understood and accepted by the archival community – a network of relationships between objects, agents and functions.[2]

It is interesting to note that the first edition of ISAD(G) assumes the physical interpretation, since it defines provenance as "the organization or individual that created, accumulated and/or maintained and used documents in the conduct of personal or corporate activity" – that is, provenance is an agent (ICA, 1994, p. 1). The first edition of ISAAR(CPF) provides the same definition of provenance (ICA, 1996, p. 1). It is only later, in the second edition of ISAD(G), that provenance becomes "the relationship between records and organizations or individuals" – that is, provenance is interpreted as a relationship rather than an agent (ICA, 2000, p. 11). However, the relationship is assumed to be singular whereas it will become plural in the subsequent documents published by ICA. Also, there is no mention of provenance as a connection between records and functions, a concept that will be introduced only in ISDF, as shown in the opening paragraph of this essay.

---

[1] Transl: "Aggregate all different records in fonds, that is, group all the documents coming from the same body, institution, family or individual, and set the different fonds according to a certain order." Charles Marie Tanneguy Comte Duchâtel, "Instructions pour la mise en ordre et le classement des archives départementales et communales," Paris le 24 avril 1841, in *Lois, Instructions et Règlements Relatifs aux Archives Départementales, Communales et Hospitalières* (Paris: H. Champion, 1884), 17.

[2] Hereafter the term "network" is used in its broader meaning as an interconnected or interrelated group of entities.

In recent years, the meaning of provenance has been investigated further, and new perspectives have been proposed:

> The similar notions of societal, parallel, and community provenance have also been advanced. They reflect an increasing awareness of the impact of various societal conditions on records creators and record creation processes at any given time and place across the records' history. [...] Some archivists have broadened the concept of provenance to include the actions of archivists and users of archives as formative influences on the creation of the records. (Nesmith, 2015, p. 286-287)

In particular, Tom Nesmith has provided a definition of provenance that, while giving rise to some issues due its very broad scope, may provide a basis for a broadened multidisciplinary perspective on provenance:

> The provenance of a given record or body of records consists of the social and technical processes of the records' inscription, transmission, contextualization, and interpretation which account for its existence, characteristics, and continuing history (1999, p. 146).

In short, archival provenance is a complex concept, the sum of different factors that altogether trace archival records back to their creation and forwards through their management and use. It is, therefore, a fundamental notion for interpreting and understanding archival objects. However, new technologies have further challenged the idea of provenance, asking for its refinement and re-interpretation. The following sections will illustrate the role of provenance in archival functions and its transformation as determined by new technologies.

The role of provenance in archival functions

It is not surprising that provenance plays a major role in different archival functions, due to its multi-dimensional nature. It plays a key role in a fundamental dimension of archival objects, that is, trust associated with them. This is especially true in the digital environment, where objects tend to float in a cyberspace with little or no context, which is great for re-using, re-purposing and re-mixing activities, yet impoverishes the objects by depriving them of their connoting qualities. This is a critical issue when we consider that such qualities are – either implicitly or explicitly – the base upon which trust is created and managed. We have moved from a physical world where documentary objects are artifacts occupying some space, to a virtual environment where objects form a vaporous nebula that we can hardly fix on the traditional axes of the Euclidean space. We need a new topology, a new way to interpret the objects of our hybrid world where virtual and real mix and overlap. As Luciano Floridi has pointed out, in the digital world location and presence are decoupled. We may be digitally present in a particular corner of the infosphere, yet our physical location may be undetermined (Floridi, 2017, p. 123-125). This holds true also for the objects and actions that belong to our space of real/virtual existence, including records and archival functions. It is a major disruption. We do not just create digital environments – we inhabit them, as spaces for social action, so we are getting to a point when we may wonder what the real thing is and what makes up its context, which is crucial for provenance.

Trust

Provenance is a crucial factor of evaluation when assessing the credibility of records on the Internet, therefore it needs to be investigated in order to shed light on the nature and the dynamics of the relationship between records and trust. The latter is a key concept of archival discipline. However, like provenance, it is a multi-faceted and cross-domain concept:

- trust is about voluntary vulnerability, in that it is based on a voluntary reliance on someone or something that may cause harm; ergo
- trust is about risk management. In fact, risk can be defined as a deviation – either positive or negative – from the expected (ISO 31000:2009, p. 1). Since trust "falls between hope and certainty," (Dietz, Gillespie, Chao, 2010, p. 13) it requires balancing confidence and control, that is, managing uncertainty, which is the essence of risk management;
- trust is a process, since the development of trust in systems as well as in people is informed by experience. Trust is built, shaped and assessed by applying known patterns to unknown situations. Therefore, trust changes over time, according to both the ever-changing factors that affect it, and people and systems' reaction to such changes;
- trust is contextual, because different systems for trust development and assessment are required for different contexts. Tools, agents, procedures, techniques and values vary according to the context; therefore,
- trust is a cultural thing. The parameters of trust in one cultural context may be very different from those of another context (Ferrin, Gillespie, 2010, p. 42-86). These parameters must be clearly identified and understood if cross-cultural trust – like what is needed on the Internet – is to be achieved;
- trust is an economic asset. In general, information has become a commodity with economic value. As a matter of fact, when exchanging information we exchange something that we consider valuable. Trust is the framework that allows such value to thrive and be exchanged.[3] However, the commodification of data – which includes sale of personal information and other datasets as well as mash-ups of data, which in turn leads to creation of new data and value – is eroding trust and consequently the value of information. This is a crucial issue in the era of open data and big data.

Like provenance, trust is a complex concept, this is the reason why it is not simple to deal with it when it comes to records. In fact, records provide evidence of our actions and thoughts, and they allow us to communicate across space and time. Such communication is deeply based on trust, to the point where trust is embedded into records. Records carry tokens of trust: signatures, seals, special signs, the documentary form itself, they all convey trust, not to mention the content, including wording and phrasing. Trust is involved in the transmission process too, since we place a certain level of confidence in the channel, the medium and the transmission service, including any associated agent and technology. (Duranti, 1998; MacNeil, 2000; Yeo, 2013, p. 214-234). The digital environment is no different, rather, it is much more complex. Digital technologies allow us to easily create, use and store documents on the Internet, where they can be de- and re-contextualized with little attention to their authenticity. Users have little control over how and where documents are stored in the Cloud, who has control and jurisdiction, who can access them, or how secure they are. In short, trust is at stake:

in the digital realm we can no longer trust documents using the same approaches and tools of the past. Therefore, provenance plays a major role here, since it is one of the crucial factors that support trust. That is why we need methods, tools and metrics – along with a solid theory – to govern provenance and support the evaluation of reliability of digital objects on the basis of information on their provenance. Prior to the digital era, archival materials were trusted because of their characteristics – as we highlighted above – or their placement within a trusted repository, i.e., an archives, with preservation, access and use of documentary objects taking place in an environment or according to processes that were considered trustable. The digital environment has corrupted such belief. The challenge today is to do something similar to what has been done with markup languages: making explicit what is implicit. Archivists and records managers need to retain control of provenance and make it explicit, so that users are aware of the quality of the objects and trust them accordingly. The challenge is to find models, methods and tools to achieve this aim, solid enough to meet scientific criteria, yet easy enough to be managed by users.

Preservation

Preservation, including digital preservation, is about keeping objects together with the context that provides meaning to them, that is, the complex network of relationships – along with the system of their meanings – in which archival objects have been created, managed and used. Provenance is highly relevant in identifying and determining such context. Consequently, it is key to determining the identity of the objects targeted for preservation, because any definition of provenance, be it narrow or broad, will address at least creation and custodial history (i.e., the chain of agents that held the materials, together with related facts and events). In addition, the provenance of digital objects is itself a digital object that requires preservation. Therefore, provenance, and provenance of provenance are fundamental aspects in any preservation model, theory and practice.[4]

Access and use

Access and preservation are two sides of the same coin. In fact, archival materials are preserved in order to make them available for use. However, "[i]n order to use a record, it must be accessible," (Kozak, 2015, p. 1) which means that policies and procedures should be designed and put in place to serve users' information needs. Provenance plays a role when accessing archival materials, since it is one of the key access points: the names of either the creator or the institution holding the archival materials are among the most common elements used in archival queries. Since provenance is more and more a complex network of relationships – if not a confused

---

[3] As Sissela Bok puts it, "[w]hatever matters to human beings, trust is the atmosphere in which it thrives." Sissela Bok, *Lying: Moral Choice in Public and Private Life* (New York, NY: Vintage Books, 1999), 31n.

[4] Significantly, the OAIS (Open Archival Information System) model – the reference model for preservation adopted worldwide – requires that any object targeted for preservation must be accompanied not only with some Representation Information providing additional, higher-level meaning to the object, but also with some Provenance Information describing the object's history (i.e., origins or source, custodial history, changes, etc.). Provenance Information is in turn a digital object. As such, it must be accompanied with some Representation Information and some Provenance Information that will document the history of the Provenance Information. Such a recursive approach creates a complex network of Information Objects that need to be managed and preserved altogether in order to provide the proper context to the objects targeted for preservation, and to support their preservation over the long term. See *ISO 14721:2012 Space Data and Information Transfer Systems: Open Archival Information System (OAIS): Reference Model* (Geneva: International Organization for Standardization, 2012).

tangle – it becomes important to allow users to understand such complexity without overwhelming them with a large mass of information. Archivists are mediators, thus they are responsible for promoting access actively and providing a perspective that puts the archival materials in context. Archival representations of provenance in the form of descriptive finding aids are a major part of this perspective. Therefore, provenance imbues the mediation tools and affects access. This is why it should be investigated thoroughly in relation to users' needs.

Appraisal

Appraisal is the process of assessing the value of records for the purpose of determining the length and conditions of their preservation.[5] According to a widespread approach known as macro-appraisal, this archival function should be based on "extensive research by archivists into institutional functionality, organizational structures and work-place cultures, recordkeeping systems, information workflows, recording media and recording technologies, and into changes in all these across space and time" (Cook, 2005, p. 103). Provenance covers several of these dimensions, once we assume that it is more than just origination, being rather a network of relationships between objects, agents and functions, so that it can be interpreted in such a broad way to cover even the social dimension.[6] As a consequence, any new understanding of the concept of provenance has a direct impact on appraisal methods and principles.

Arrangement and description

Arrangement and description of archival materials require identification and description of both the creators and the chain of custody of materials. When arranging, provenance is the first clue enabling archivists to trace archival materials back to their origins, identify different bodies of materials, and get to a tentative grouping. When describing, the complexity of provenance may affect the representation of the archival materials. This is indeed more true in the digital realm, where new visualization tools and information models allow for greater freedom when designing archival descriptions. At the same time, representation models affect the ways that provenance is understood and represented in archival descriptions, because they highlight certain features while hiding or obfuscating others. Moreover, materials on the Internet are not only dispersed but are also mixed and re-used to a point that it is often difficult to trace provenance. In short,

---

[5] This is the traditional and consolidated definition of appraisal. "Appraisal [is t]he process of determining the retention period of records" according to ICA. See International Council on Archives, *ISAD(G) 2nd edition*, 10. Similar definitions can be found on the most authoritative sources: the Multilingual Archival Terminology (MAT) published online by the International Council on Archives defines appraisal as "[t]he process of identifying materials offered to an archives that have sufficient value to be accessioned." See ICA MAT, accessed October 6, 2017, http://www.ciscra.org/mat/mat/term/47. The Glossary of Archival and Records Terminology adopted by the Society of American Archivists provides this same definition along with a similar one: "[T]he process of determining the length of time records should be retained, based on legal requirements and on their current and potential usefulness." See Richard Peirce-Moses, *A Glossary of Archival and Records Terminology* (Chicago: SAA, 2005), 22. However, in recent years new definitions have appeared: the so-called ICA Req defined appraisal as "[t]he process of evaluating business activities to determine which records need to be captured and how long the records need to be kept, to meet business needs, the requirements of organisational accountability and community expectations." See International Council on Archives, *Principles and Functional Requirements for Records in Electronic Office Environments* (Paris: ICA, 2008), 73. A more disruptive definition appeared in 2017: ISO 15489-1 defines appraisal as "[t]he process of evaluating business activities to determine which records need to be created and captured and how long the records need to be kept." See *ISO 15489-1 Information and documentation: Records >*

provenance is a crucial dimension of any arrangement and description process. Also, with a growing number of records being created and preserved using Cloud technology, there is a need to consider how to undertake their arrangement and description in the Cloud. To this end, a research project has been set up in the broader context of the InterPARES Trust, a "multi-national, interdisciplinary research project exploring issues concerning digital records and data entrusted to the Internet,"[7] launched in 2013 and led by the University of British Columbia. The specific project, titled "Arrangement and Description in the Cloud," investigates how the Cloud environment is going to affect arrangement and description theory and practice.[8] Only a preliminary analysis of the problem has been conducted so far, yet some interesting observations have emerged from such analysis.[9]

a. Archives are beginning to implement and develop services that capture records from Cloud-based services such as providers of email and social media services. Generally, a software application will connect to the Cloud service using whatever method the service provider specifies. In the case of social media, the capture tool connection is likely to interact with Application Programming Interfaces (APIs) that operate according to rules defined by the service provider. Using tools such as ePADD, Social Feed Manager, ArchiveSocial (itself a Cloud service), Thinkup, or the Twitter Archiving Google Sheet, a record is fixed in place by a Cloud provider such as Google, or social media services like Twitter, Instagram, Flickr, and Facebook.[10] Such tools collect vast amounts of metadata of potential value in tracking not only the origin and use of a particular tweet, but also regarding how the archivist shaped the collection. However, tweet-specific metadata may be stored in a way that makes them transparent to the other applications. For example, the Social Feed Manager stores metadata in WARC files,[11] which means that whatever provenance or other metadata exist for a tweet is kept in a JSON format as part of a WARC file. In other words, such provenance metadata is not immediately known to the database-driven parts of the application. In addition, resources that are referenced in the tweet, either as embedded or external content (e.g., images, videos and webpages), are captured in the WARC file. In theory, many types of metadata at all levels could be controlled in an archival descriptive system. However, key questions, such as which metadata to extract and ingest into the archival management system, remain to be investigated.

5   *management: Part 1: Concepts and principles* (Geneva: International Organization for Standardization, 2016), 10. However, this brand-new definition is rooted in a specific geo-cultural context and is not agreed by the archival community at large, so we will refer here to the consolidated definition of appraisal.

6   *Vide supra.*

7   *InterPARES Trust*, accessed October 6, 2017, https://interparestrust.org/trust.

8   The research team includes Giovanni Michetti, Richard Peirce-Moses, Chris Prom and Kat Timms.

9   The following three paragraphs are drawn with changes from Christopher Prom, Giovanni Michetti, Katherine Timms, Andrea Tarnawski and Richard Peirce-Moses, "Archival Arrangement and Description in the Cloud: A Preliminary Analysis," in *Proceedings of XXI Archival Science Colloquium, Marburg, 8 June 2016* (Marburg, DE: Archivschule, in press).

10 EPADD, https://library.stanford.edu/projects/epadd; Social Feed Manager (SFM), http://gwu-libraries. github.io/sfm-ui/; ArchiveSocial http://archivesocial.com; ThinkUp, https://github.com/ThinkUpLLC/ ThinkUp; Twitter Archiving Google Sheet (TAGS), https://tags.hawksey.info. All websites accessed October 6, 2017.

11 Web ARChive (WARC) is an ISO standard for web archiving. This format aggregates multiple digital resources into a single file together with related information. See *ISO 28500:2009 Information and Documentation: WARC File Format* (Geneva: International Organization for Standardization, 2009).

b. Several studies note that as technology develops, new value can be assigned to records; this is particularly true with Cloud services. For example, Instagram is used as both a "storage box" of personal photos and a space to share information about users' identity and activities.[12] Should the archival management system capture and preserve the profile in place at the moment of creation or transmission of each record? Additional complexities arise when new people enter the picture. The collaborative nature of social media platforms encourages the creation of new records (or new representations of existing records) via linkages, embedded content and comments. "Likes," tags, and participation by others on photos add new value to those possessions, but such metadata can easily become obscured in the interface, if not trapped in the application where it is recorded. The additional information added by others might be considered as context-of-creation metadata (in the case of collaborative environments such as Google Drive) or context-of-use metadata, such as "likes" and "shares" in a social media platform. Both forms of context suggest that archival systems will need a method to represent the role that a particular user played in modifying or adding to the core record, that is to say, the original "creation" developed by the original "author," "creator," or "collector" of a particular work (Bak, Hill, 2015, p. 101-161). Archival descriptive records might somehow catch and fix these new associations as some representation of provenance.[13]

Context is and has always been a fluid entity in time, that is, it changes as time passes by. What is new today is that context has become a fluid entity in space, that is, it changes as we look at it from a different perspective. For example, a document stored in Google Drive or a similar Cloud-storage service may be represented as belonging to one folder for the original creator and a different folder for a contributor provided permission to update the document. Given the collaborative nature of these tools, it appears that in general the same document belongs to different folders according to the agent – be it an individual or a system – that interacts with the document.[14] Similarly, social media postings appear at a particular point in a stream of posts. The specific stream is produced by the interaction of object metadata with user preferences and choices, and these of course vary for different users at different times; as users comment on or annotate that record, evidence about its use accrues alongside the original post. The consequential question is whether the standards and tools available to archivists will allow them to preserve both the records and the complex relationships reflecting their creation and use, which represent a major part of their context. A preliminary question should be whether archivists agree that such network of relationships

---

[12] The term "storage box" is used by Odom, Sellen, Harper and Thereska to illustrate how causal users may treat networked environments as a place to make digital materials accessible across different physical places or using it as an alternative place of storage for backup purposes. See William Odom et al., "Lost in Translation: Understanding the Possession of Digital Things in the Cloud," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Austin, 2012* (New York, NY: ACM, 2012), 781-790.

[13] New representations of provenance as a more complete set of information about actions taken in the origination and subsequent handling of a digital object can be represented in records complying with the requirements of the PROV Ontology. See Paolo Missier, Khalid Belhajjame and James Cheney, "The W3C PROV Family of Specifications for Modelling Provenance Metadata," in *Proceedings of the 16th International Conference on Extending Database Technology, Austin, 2012* (New York, NY: ACM, 2013), 773.

[14] Please note that we are not referring to the case in which a document is assigned to different folders for records management purposes. We are referring to the fact that a specific document gets a different context according to the user that interacts with it.

needs to be preserved. If so, what can be done to help them implement a cohesive set of archival services that are suitable to the Cloud-based environment in which many people live their digital lives? Should archivists stick to a static, single perspective framing data and metadata once it crosses the archival threshold, or should they adopt a more flexible approach where different perspectives may coexist? What metadata should be retained? For what purposes?

Furthermore, how much metadata is enough? In the digital environment, metadata associated with or embedded into records may provide relevant information on the provenance of either the records themselves or the systems in which they reside. However, if the scope of provenance is broadened to include societal provenance,[15] the list of sources where to get metadata needs to be extended to include materials documenting aspects of both the society at large and the specific communities in which the records have been created, managed and used.

Linked Data

The most promising model for describing digital resources is RDF (Resource Description Framework).[16] Its very simple design is based on the notion of a triple, that is, a statement consisting of a subject, a predicate, and an object, describing some elemental aspects of a resource. RDF is a fundamental component of the Semantic Web architecture, since it allows – along with other Web technologies – to publish and interlink structured data that can support semantic queries, i.e., queries that enable the retrieval of both explicit and implicit information.[17] Data published on the Web according to this architecture are called Linked Data.[18] Ontologies complement and enhance the power of Linked Data, as they are formal specifications of a shared conceptualization, and act as a cornerstone of defining a knowledge domain. Tim Berners-Lee established four simple rules for creating Linked Data:

“1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
4. Include links to other URIs so that they can discover more things (Berners-Lee, 2006)”.

It is interesting to note that Linked Data seem to be a perfect fit for the nebula of data objects mentioned above: statements can be linked to other statements,

---

[15] Societal provenance is a term used to mean provenance in the broader sociocultural dimension. Records creation, management, use and preservation are sociocultural phenomena. Therefore, provenance may be interpreted taking into account the sociocultural dimension as the context in which all actions take place.

[16] For more information on RDF, see https://www.w3.org/RDF/.

[17] The triples describe resources, so they may be interpreted as metadata, that is, data about data. However, it is important to highlight that being metadata is not an ontological property, since there is no such thing as metadata per se. Some data are called metadata, because a special value is assigned to them – they are recognized as conveying information on some specific dimension considered as being relevant in a given context. For example, dates are usually considered metadata, because of the relevance of the temporal dimension. At the same time, dates are data, because they are usually embedded into documents, that is, they are integral part of the datum. There is no antithesis nor contradiction – everything is data. Sometimes it is called metadata to highlight its special value.

[18] RDF is a data architecture, while Linked Data is a way of publishing RDF data.

which may lead to an ever-expanding set of statements taking the form of a graph. Therefore, it is not surprising that Linked Data are disseminated on the Web more and more, and both archivists and records managers are slowly following this trend, creating and distributing information in the form of Linked Data, thus changing system designs and descriptive practices. However, the archival community has not yet developed an ontology modelling and representing provenance, whereas the data science community has already created its own ontology for representing entities and relationships with respect to the origin and provenance: the PROV ontology defines provenance as "information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness." (World Wide Web Consortium, 2013). The basic model of the PROV ontology is simple and recursive, allowing for great complexity and expressiveness. Its core concepts are Entities, Agents, and Activities (Figure 1).[19] Call them Objects, Agents and Functions, and the picture in Figure 1 will make perfect sense in the archival domain.[20]
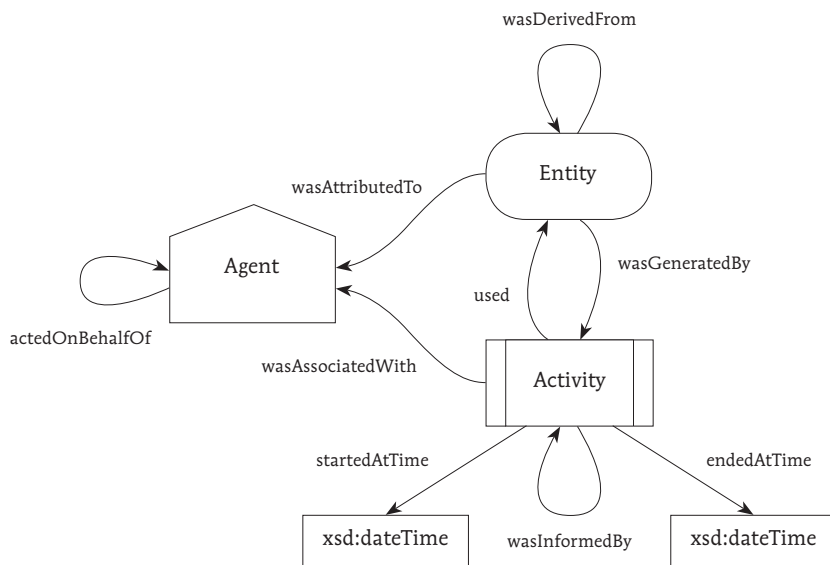
Figure 1. The basic classes and properties of the PROV ontology.

[19] The PROV ontology defines the classes Entity, Activity and Agent, and link such classes through properties, as shown in Figure 1. An Entity is a "physical, digital, conceptual, or other kind of thing with some fixed aspects; [it] may be real or imaginary." An Activity is "something that occurs over a period of time and acts upon or with entities; it may include consuming, processing, transforming, modifying, relocating, using, or generating entities." An Agent is "something that bears some form of responsibility for an activity taking place, for the existence of an entity, or for another agent's activity." See World Wide Web Consortium, *PROV-O: The PROV Ontology*, W3C Recommendation 30 April 2013, eds. Timothy Lebo, Satya Sahoo and Deborah McGuinness, accessed October 6, 2017, https://www.w3.org/TR/2013/REC-prov-o-20130430/.

[20] The diagram included in ISDF is not much different. See International Council on Archives, *ISDF: International Standard for Describing Functions* (Paris: ICA, 2007), 36.

The PROV ontology focuses on lineage, that is, on data's origins and history, in order to provide a tool that may help tracing the objects back to their creation, which is a critical issue for the verification of the process used to generate the data. Provenance, particularly in the archival domain, is a broader concept that serves to identify and possibly document any input, entity, system and process that has affected data.[21] Therefore, while the PROV ontology may be a good starting point, there remains much space for discussion on how to best model archival provenance. What is quite clear though is that we are very likely to adopt the Linked Data paradigm. Therefore, it is worth highlighting some pros and cons of Linked Data, in order to understand the effects on provenance.

Linked Data: benefits

The benefits of Linked Data are quite evident: due to their characteristics (i.e., the compliance with the requirements established by Tim Berners-Lee) they are inherently shareble, extensible and re-usable.[22] Resources – be they physical or conceptual – are identified by language-agnostic URIs and connected through properties that can be identified by a URI as well. This leads to the creation of a gigantic network: this mechanism allows anyone not only to add further descriptions (i.e., statements) that increase the size and the density of the original network of relationships; but also, to establish new connections among different, separated networks. Once a connection is made, the two networks become a whole, that is, an enlarged and enriched network of relationships, which eventually leads to a Giant Global Graph, as Tim Berners-Lee called it (Berners-Lee, 2007). In addition, semantic enrichment – that is, the process of adding layers of metadata to content – allows computers to process data and make sense of it, so that they can find, filter, and connect information.

Linked Data in the archival domain allow for the creation of new pathways not only for exploring archival descriptions, but also for accessing the resources themselves. In fact, on the one hand Linked Data makes it possible to explore the complex web of relationships that make up and define the context in which objects are placed, and that has a fundamental value in Archival Science. In this sense, we could say that Linked Data increase the possibilities for exploring the metalevel (that is, the descriptive dimension). On the other hand, the fragmentation of information operated by Linked Data creates new and numerous points of direct access to resources. In theory, Linked Data are an unlimited source of access points.[23] This is perfectly in line with the non-specialist research practices suggested by search engines: more and more users get to the archival sources through Google and other search engines, with an approach that favors the process of disintermediation

---

[21] Some authors distinguish between different types of Provenance, such as Why-Provenance, How-Provenance, Where-Provenance, and Workflow-Provenance. See for example James Cheney, Laura Chiticariu and Wang-Chiew Tan, "Provenance in Databases: Why, How, and Where," *Foundations and Trends in Databases 1* (2009): 379-474; Peter Buneman, Sanjeev Khanna and Wang Chiew Tan, "Why and Where: A Characterization of Data Provenance," in *Proceedings of the 8th International Conference on Database Theory, London, 2001,* eds. Jan Van den Bussche and VictorVianu (London, UK: Springer, 2001), 316-330; Susan B. Davidson and Juliana Freire, "Provenance and Scientific Workflows: Challenges and Opportunities," in *Proceedings of the SIGMOD International Conference on Management of Data, Vancouver, 2008,* ed. Jason Tsong-Li Wang (New York, NY: ACM, 2008), 1345-1350. According to this perspective, lineage is a type of Provenance.

[22] This is especially true for Linked Open Data (LOD), that is, "Linked Data which is released under an open licence, which does not impede its reuse for free." Berners-Lee, *Linked Data*.

[23] Even with an advanced descriptive model like EAD we must go through a tag like <controlaccess> to identify the access points.

between source and user. However, search is not the only strategy for exploring archives-users may discover archival materials by moving around in the digital environment. Search is great if we know what we are looking for, but discovery reveals what we did not know existed, it generates new relationships. Linked Data support discovery thanks to their intrinsic nature: the underlying graph is not only a data architecture, but also a network of nodes that can be used as a path to explore freely the vastitude of online resources.

Linked Data: risks

Unfortunately, the granularity of Linked Data runs counter to current descriptive practices, characterized by the abundant use of free text in archival descriptions, a condition that severely limits the possibilities for interoperability and perpetuates the isolation of archival data, preventing integration with other types of data. This is an inherent limitation of the most prevalent forms of archival representation (inventories and guides in particular), which makes the adoption – rather, the exploitation – of the RDF model difficult. As a matter of fact, all archival descriptive models, including EAD, favor the narrative character of the finding aid. As noted many years ago by Elizabeth Yakel, "the concentration on the finding aid as document rather than as one of many potential representations of discrete data elements has led to problems of reusing archival data archival across the archival continuum and problems in the development of true collection management systems for archives." (Yakel, 2003, p. 18).

Trying to move a step further, the International Council on Archives initiated years ago a process of revision of its standards for archival description. This initiative has led to the publication of a new conceptual model in September 2016, clearly and explicitly driven by the RDF data architecture (ICA, 2016). Therefore, this model takes into account the technological developments of recent years, and builds on the idea of graph as the ideal architecture for conveying information on the context: "Modelling description as a graph accommodates the single, fonds-based, multilevel description modelled in ISAD(G), but also enables addressing the more expansive understanding of provenance described above." (ICA, 2016) ICA intends to move archival description from a multi-level to a multi-dimensional approach: "The multidimensional model […] sees the fonds existing in a broader context, in relation to other fonds. In a multidimensional approach to description, the Records and Sets of Records, their interrelations with one another, their interrelations with Agents, Functions, Activities, Mandates, etc., and each of these with one another, are represented as a network within which individual fonds are situated." (ICA, 2016)

This initiative has adopted the key words in current information architecture: graph, multi-dimensionality, networks of interrelations. However, this document raised some relevant objections in the archival community, with regard to different aspects.[24] In particular, InterPARES Trust, a large community of hundreds of

---

[24] Some critical comments have been posted to both the ICA mailing list devoted to this initiative (ICA-EGAD-RiC Mailing List, http://lists.village.virginia.edu/mailman/options/ica-egad-ric) and the ICA mailing list (ICA Mailing List, http://www.ica.org/en/ica-list-serv). Chris Hurley has published on his blog a dense critique on RiC opening his post with a short yet effective consideration: "RiC is a conceptual model in search of a concept." See Chris Hurley, "RiC at Riga," *Chris Hurley's Stuff*, August 2017, http://www.descriptionguy.com/images/WEBSITE/ric_at_riga.pdf. William Maher, in his role of Chair of the ICA Section on University and Research Institution Archives, has raised some reasonable and thoughtful >

researchers from all over the world, laid down a set of critical comments about the fairness and transparency of the process, the methodology adopted for developing the model, and the model itself. The concluding statements of the document submitted by InterPARES Trust are quite explicit:

> In short, we find that RiC-CM is weak as a model, in that it neither defines the structures it uses (entity, property, relation) nor provides a rationale for their use. A conceptual model should identify and define the fundamental *bricks* used to build the model. […] Ultimately, the document fails to adequately address a model for discovery of archival resources, a model that accommodates multiple users and uses. […] EGAD and ICA should re-start the development process on a new, transparent and fair basis […]. (InterPARES Trust, 2016)

It will be interesting to see whether and how these concerns will be addressed in the future, and – in case – where this will lead the concept of provenance. As noted before, in the past twenty years the International Council on Archives has changed its approach to provenance a few times, interpreting it first as an agent, then as a single relationship, later as a set of relationships, and now as a multi-dimensional concept. Therefore, there is some reason to believe this is neither the perfect solution nor the final step.

Another issue to consider when dealing with Linked Data is expressed outright by Hay Kranen in his blog: "Linked data is all nice and dandy, but if your SPARQL endpoint is only up 50% of the time and it takes a minute to do a query, how do you suppose a developer builds a stable app on top of it?" (Kranen, 2014) The post dates back to 2014, but it still holds true: keeping an endpoint up can be challenging. In a comment to the same post, Marcus Smith noted: "It's almost become an in-joke that six simultaneous users of a SPARQL endpoint constitutes a DDOS attack." In fairness, it should be recognized that endpoints and triple-store technologies are young, so it is likely that the situation will improve in the course of time.

The fact that the Semantic Web technologies are rather difficult to implement and require high skills is another issue to consider when dealing with Linked Data. However, this too is a problem related to technologies that are still not completely mature: probably it still needs some time before both technologies and skills become less esoteric.

Most of all, the fundamental problem of Linked Data lies in their very structure. The critical problem is the graph. As Bowker and Star note, "[e]ach standard and each category valorizes some point of view and silences another. This is not inherently a bad thing – indeed it is inescapable. But it is an ethical choice, and as such it is dangerous – not bad, but dangerous." (Bowker, Leigh Star, 2000, p. 5-6) We need to

---

> doubts about RiC in relation to archivists' missions and mandates. See William J. Maher, *ICA-SUV 2017 Conference Summary*, accessed October 6, 2017, https://icasuvblog.wordpress.com/2017/09/13/ica-suv-2017-conference-summary/. RiC describes as much as seventy-three "potential record-to-record relations". Instead of "seeking an exhaustive list of every relation that might exist between two records," Ross Spencer has taken a different approach and has outlined in his essay eight relations only. See Ross Spencer, "Binary trees? Automatically identifying the links between born-digital records," *Archives and Manuscripts* 45 (2017): 77-99.

understand the meanings and biases hidden in our professional tools, practices and theories. "Recognizing the presence of an underlying paradigm and understanding the values it conveys is not difficult when we deal with concepts, principles and categories, while it may be tricky when we deal with technical, apparently neutral standards. In fact, different technologies may rely on different philosophies." (Michetti, 2015, p. 155) So far, archivists and records managers have focused on the documentary object as a whole. RDF and Linked Data are almost a Copernican revolution, because they rely on information atoms that – in theory – can be aggregated and manipulated at will. This is the perfect solution for those like Greg Bak who advocate an item-level thinking (Bak, 2012). However, the adoption of XML, RDF, Linked Open Data and other technologies is more than a technical option: it is rather the choice of a specific knowledge paradigm, not at all neutral. In the case of Linked Data, the graph is not only the symbolic representation of the network of relationships among the entities that make up the archival description. It is also the form taken by data, the structure that houses the descriptions, the container that gives shape to our vision of the world. To paraphrase Bowker and Star, there is nothing wrong with that. However, we need to understand the profound significance of this approach.

The graph offers many advantages, but its strength – that is, the potential to create a network of connections that can be expanded indefinitely – can prove to be a limit. For example, if we consider EAD, it is evident that its limit resides in its design, that is, in thinking and designing an archival description as a document. As a matter of fact, EAD provides a digital replica of the paper object. However, it is also true that this approach has still some reasons, when we recognize that archival description is an autonomous work. In fact, in addition to practical and operational purposes, archival description has also a fundamental function of mediation between sources and users, and supports the authenticity of the sources. In a graph, it can be difficult to recognize the boundaries of a given archival description. With Linked Data, Anyone can say Anything about Anything[25]: once we accept this so-called Principle of the triple A, links explode – that's the beauty of Linked Data –, boundaries disappear and users can access directly from anywhere in the graph. In a sense, this is a profound form of disintermediation that is destined to grow as visualizations techniques and strategies occupy the archival space, dominated so far by written word, narrative and hierarchical diagrams. The complex network of relationships underlying – rather, making up – an archive can be now be represented in a myriad of ways. This is not a criticism of Linked Data: the graph paradigm is indeed a promising data architecture. This is rather an exploration of the possible limits and dangers of this paradigm. In short, archivists should investigate this transformation process that is slowly moving archival description in a direction that leads to bibliographic description: high fragmentation of information, and reduction of the narrative dimension.

---

[25] "To facilitate operation at Internet scale, RDF is an open-world framework that allows anyone to say anything about anything. In general, it is not assumed that all information about any topic is available. A consequence of this is that RDF cannot prevent anyone from making nonsensical or inconsistent assertions, and applications that build upon RDF must find ways to deal with conflicting sources of information." World Wide Web Consortium, *Resource Description Framework (RDF): Concepts and Abstract Data Model*, W3C Working Draft 29 August 2002, eds. Graham Klyne and Jeremy Carroll, accessed October 6, 2017, https://www.w3.org/TR/2002/WD-rdf-concepts-20020829/#xtocid48014.

Finally, it should be noted that the effects of the principle of triple A are multiplied when added to the Open World Assumption (OWA). Roughly speaking, this assumption states that that the absence of a statement does not imply a declaration on the absence (for example, the absence of date of birth does not mean that the person is not born).[26] Under these conditions, what value should be attributed to the statements (i.e., the triples)? The question is not trivial and indeed takes us back to issues such as source of authority and technical expertise, which have a deep connection with provenance and thus should be taken into account when designing new models for archival description. Strategies are needed to assess users' trust in relation to the quality of information on provenance. After all, this brings us back to the trust issue that Tim Berner-Lee already identified at the top of the Semantic Web stack (Berners-Lee, 2000).

Conclusions

As already stated and discussed, the Principle of Provenance is a pillar of Archival Science, originally intended to prevent the intermingling of documents from different origins, in order to maintain the identity of a body of records. Peter Scott challenged such a view. As a consequence, provenance in the archival domain moved from a simplistic ono-to-one relationship to a multi-dimensional approach, and started being understood as a network of relationships between objects, agents and functions. Conceptual debate pushed the boundaries of provenance further: the established orthodoxies cracked under the weight of societal, parallel and community provenance. The digital environment and new technologies have presented unpredictable challenges to the concept of provenance: not only are digital objects often the result of an aggregation of several different pieces, but it also is extremely easy to mix and re-use them, to a point where it may be very difficult to trace their provenance. Cloud Computing has complicated the picture further, due to the little control that it is possible to exercise over the Cloud service providers and their procedures. As a result, the archival functions are compromised, since objects get their meaning from their context, and provenance plays a major role in identifying and determining such context: whenever provenance is flawed, so is context, hence the overall meaning of an object. Moreover, any lack of control over provenance determines uncertainty, which in turn affects trust in digital objects, thus hindering the implementation of the top level of the Semantic Web stack designed by Tim Berners-Lee.

However, new technologies provide a solution to cope with such complexity. Resource Description Framework (RDF) and ontologies can be used to represent provenance through new standards and models in a granular and articulated way that was not conceivable before the advent of computers. Provenance is slowly taking the form of a network of triples, that is, a complex set of interrelated statements that is apparently distant from the original Principle of Provenance, yet

---

[26] The Open World Assumption codifies the informal notion that in general no single agent or observer has complete knowledge. Not surprisingly, the Semantic Web makes the Open World Assumption.

it is rooted in that idea. RDF triples can be used to express specific types of relationships and establish different connections among entities. There would be no need to agree that certain elements are integral to provenance and to reject certain others: the story could simply be told, and the model for telling it could be made sufficiently comprehensive to allow everyone to tell their stories.

Therefore, the digital environment is indeed a source of new problems, but it is also an opportunity to review and refine established practices and concepts. Probably technology is not the hardest issue. The major challenge is a change of mindset, that is, moving from a Ctrl-c Ctrl-v attitude, a trivial operation "where much provenance gets lost," (Buneman, Davidson, 2010) to a more responsible approach that could be supported by and embedded into system design. After all, there is already Privacy by Design, Quality by Design, Security by Design, and so on – the time has come for Provenance by Design.

Literature

Bak, Greg, & Amanda Hill (2015). Deseronto Dreams: Archives, Social Networking Services, and Place. *Archivaria, 79*, 1-26.

Bak, Greg (2012). Continuous Classification: Capturing Dynamic Relationships among Information Resources. *Archival Science, 12,* 287-318.

Berners-Lee, Tim (2006). *Linked Data*, July 27, 2006. Retrieved from https://www.w3.org/DesignIssues/LinkedData.html

Berners-Lee, Tim (2007). *Giant Global Graph*, November 11, 2007. Retrieved from http://dig.csail.mit.edu/breadcrumbs/node/215, archived at https://web.archive.org/web/20170423133504/http://dig.csail.mit.edu/breadcrumbs/node/215

Berners-Lee, Tim (2000). *Semantic Web on XML*, slides, 2000. Retrieved from https://www.w3.org/2000/Talks/1206-xml2k-tbl/slide1-0.html

Bok, Sissela (1999). *Lying: Moral Choice in Public and Private Life*. New York: Vintage Books.

Bowker, Geoffrey C., & Susan Leigh Star (2000). *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.

Buneman, Peter, Sanjeev Khanna, & Wang Chiew Tan (2001). Why and Where: A Characterization of Data Provenance. In Jan Van den Bussche and VictorVianu (eds.), *Proceedings of the 8th International Conference on Database Theory*, London, 2001 (pp. 316-330). London: Springer.

Buneman, Peter, & Susan B. Davidson (2010). *Data Provenance: The Foundation of Data Quality*, September 1, 2010. Retrieved from https://www.sei.cmu.edu/measurement/research/upload/Davidson.pdf

Cheney, James, Laura Chiticariu, & Wang-Chiew Tan (2009). Provenance in Databases: Why, How, and Where. *Foundations and Trends in Databases, 1*, 379-474.

Cook, Terry (2005). Macroappraisal in Theory and Practice: Origins, Characteristics, and Implementation in Canada, 1950-2000. *Archival Science, 5*, 101-161.

Davidson, Susan B., & Juliana Freire (2008). Provenance and Scientific Workflows: Challenges and Opportunities. In Jason Tsong-Li Wang (ed.), *Proceedings of the SIGMOD International Conference on Management of Data, Vancouver, 2008* (pp. 1345-1350). New York, NY: ACM.

Dietz, Graham, Nicole Gillespie, & Georgia T. Chao (2010). Unravelling the Complexities of Trust and Culture. In Mark N. K. Saunders, Denise Skinner, Graham Dietz, Nicole Gillespie and Roy J. Lewicki (eds.), *Organizational Trust: A Cultural Perspective* (pp. 3-41). Cambridge, UK: Cambridge University Press.

Duchâtel, Comte Charles Marie Tanneguy (1884). Instructions pour la mise en ordre et le classement des archives départementales et communales, Paris le 24 avril 1841. In *Lois, Instructions et Règlements Relatifs aux Archives Départementales, Communales et Hospitalières*. Paris: H. Champion.

Duranti, Luciana (1998). *Diplomatics: New Uses for an Old Science*. Lanham, MD: Scarecrow Press.

Ferrin, Donald L., & Nicole Gillespie (2010). Trust Differences Across National-societal Cultures: Much to Do, or Much Ado about Nothing? In Mark N. K. Saunders, Denise Skinner, Graham Dietz, Nicole Gillespie and Roy J. Lewicki (eds.), *Organizational Trust: A Cultural Perspective* (pp. 42-86). Cambridge, UK: Cambridge University Press.

Floridi, L. (2017). Digital's Cleaving Power and Its Consequences. *Philosophy and Technology, 30*, 123-129.

International Council on Archives (1996). *ISAAR(CPF): International Standard Archival Authority Record for Corporate Bodies, Persons and Families*. Ottawa: ICA.

International Council on Archives (1994). *ISAD(G): General International Standard Archival Description*. Ottawa: ICA.

International Council on Archives (2000). *ISAD(G): General International Standard Archival Description* (2nd ed.). Ottawa: ICA.

International Council on Archives (2007). *ISDF: International Standard for Describing Functions*. Paris: ICA.

International Council on Archives (2008). *Principles and Functional Requirements for Records in Electronic Office Environments*. Paris: ICA.

International Council on Archives (2016). *Records in Contexts: A Conceptual Model for Archival Description*, Consultation Draft v0.1, September 2016. Retrieved from http://www.ica.org/sites/default/files/RiC-CM-0.1.pdf

InterPARES Trust (2016). *Comments on "Records in Context",* December 10, 2016. Retrieved from https://interparestrustblog.files.wordpress.com/2016/12/interparestrust_commentsonric_final2.pdf.

ISO 14721:2012 Space Data and Information Transfer Systems: Open Archival Information System (OAIS): Reference Model. Geneva: International Organization for Standardization, 2012.

ISO 15489-1 Information and documentation: Records management: Part 1: Concepts and principles. Geneva: International Organization for Standardization, 2016.

ISO 28500:2009 Information and Documentation: WARC File Format. Geneva: International Organization for Standardization, 2009.

ISO 31000:2009 Risk management: Principles and Guidelines. Geneva: International Organization for Standardization, 2009.

Kozak, Greg (2015). Access/Accessibility. In Luciana Duranti and Patricia C. Franks (eds.), *Encyclopedia of Archival Science* (pp. 1-3). Lanham, MD: Rowman & Littlefield.

Kranen, Hay (2014), Linked Data and the Semantic Web: GLAM's betting on the wrong horse? *Hay Kranen's Blog*, 19 May 2014, retrieved from https://www.haykranen.nl/2014/05/19/linked-data-and-the-semantic-web-glams-betting-on-the-wrong-horse/

MacNeil, Heather (2000). *Trusting Records: Legal, Historical and Diplomatic Perspectives*. Dordrecht, NL: Kluwer Academic Publishers.

Michetti, Giovanni (2016). Provenance: An Archival Perspective. In Victoria L. Lemieux (ed.), *Building Trust in Information: Perspectives on the Frontiers of Provenance* (pp. 59-68). Cham, CH: Springer International Publishing.

Michetti, Giovanni (2015). Unneutrality in Archival Standards and Processes. In Franjo Pehar, Christian Schlögl and Christian Wolff (eds.), *Reinventing Information Science in the Networked Society: Proceedings of the 14th International Symposium on Information Science (ISI 2015),* Zadar, 2015 (pp. 144-159). Glückstadt, DE: Verlag Werner Hülsbusch.

Missier, Paolo, Khalid Belhajjame, & James Cheney (2013). The W3C PROV Family of Specifications for Modelling Provenance Metadata. In *Proceedings of the 16th International Conference on Extending Database Technology*, Austin, 2012, (pp. 773-776). New York, NY: ACM.

Muller, Samuel, Johan Adriaan Feith, & Robert Fruin (2003). *Manual for the Arrangement and Description of Archives* (2nd ed., translated by Arthur H. Leavitt). Chicago: Society of American Archivists.

Nesmith, Tom (2015). Principle of Provenance. In Luciana Duranti and Patricia C. Franks (eds.), *Encyclopedia of Archival Science* (pp. 284-288). Lanham, MD: Rowman & Littlefield.

Nesmith, Tom (1999). Still Fuzzy, but More Accurate: Some Thoughts on the 'Ghosts' of Archival Theory. *Archivaria, 47*, 136-150.

Odom, William, Abigail Sellen, Richard Harper, & Thereska Eno (2012). Lost in Translation: Understanding the Possession of Digital Things in the Cloud. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Austin, 2012 (pp. 781-790). New York, NY: ACM.

Prom, Christopher J., Giovanni Michetti, Katherine Timms, Andrea Tarnawsky, & Richard Pearce-Moses (2016). Archival Arrangement and Description in the Cloud: A Preliminary Analysis. *In Proceedings of XXI Archival Science Colloquium: Marburg 8 June 2016*. Marburg, DE: Archivschule, in press.

Scott, Peter J. (1966). The Record Group Concept: A Case for Abandonment. *The American Archivist, 29*, 493-504.

Spencer, Ross (2017). Binary trees? Automatically identifying the links between born-digital records. *Archives and Manuscripts, 45*(2), 77-99.

World Wide Web Consortium (2002). *Resource Description Framework (RDF): Concepts and Abstract Data Model*, W3C Working Draft 29 August 2002, edited by Graham Klyne and Jeremy Carroll. Retrieved from https://www.w3.org/TR/2002/WD-rdf-concepts-20020829/#xtocid48014

World Wide Web Consortium (2013a). *PROV-O: The PROV Ontology*, W3C Recommendation 30 April 2013, edited by Timothy Lebo, Satya Sahoo and Deborah McGuinness. Retrieved from https://www.w3.org/TR/2013/REC-prov-o-20130430/

World Wide Web Consortium (2013b). *PROV-Overview: An Overview of the PROV Family of Documents*, W3C Working Group Note 30 April 2013, edited by Paul Groth and Luc Moreau. Retrieved from https://www.w3.org/TR/2013/NOTE-prov-overview-20130430/

Yakel, Elizabeth (2003). Archival Representation. *Archival Science, 3,* 1-25.

Yeo, G. (2013). Trust and Context in Cyberspace. *Archives and Records: The Journal of the Archives and Records Association, 34*, 214-234.

Websites (accessed October 6, 2017)

ArchiveSocial. http://archivesocial.com
EPADD. https://library.stanford.edu/projects/epadd
ICA Mailing List. http://www.ica.org/en/ica-list-serv
ICA-EGAD-RiC Mailing List. http://lists.village.virginia.edu/mailman/options/ica-egad-ric
InterPARES Trust. https://interparestrust.org/trust
Multilingual Archival Terminology (MAT). http://www.ciscra.org/mat/
Resource Description Framework (RDF). https://www.w3.org/RDF/
Social Feed Manager (SFM). http://gwu-libraries.github.io/sfm-ui/
ThinkUp. https://github.com/ThinkUpLLC/ThinkUp
Twitter Archiving Google Sheet (TAGS). https://tags.hawksey.info

FRANS SMIT

Records, Hyperobjects and Authenticity

Overview

In this essay I argue that authenticity will remain an essential qualification of records in a digital environment. Authentication, is, and will stay, dependent on the interaction between three entities: the authenticator, the authenticated and the informed. This interaction however, will undergo fundamental changes.

The reason for this is that the nature of records has changed dramatically in a digital environment. They have become so vast and so distributed that we might view them as part of a Hyperobject, along the line of reasoning of the ecological philosopher Timothy Morton. From that new perspective I will explore the concept of the Hyperobject to grasp this nature. I will also examine how authentication of a hyperobject might work. The essay will conclude with some general observations on what is new and what is not; and on ethical challenges considering the authentication of this hyperobject.

In the essay a lot of topics, questions and deliberate choices are addressed. Therefore, it can only depict a snapshot of a line of reasoning that should be elaborated.[1]

Authenticity: a word of ominous import

Writing about authenticity is a hazardous enterprise. There are moral philosophers who try to convince us that we are currently "living in an Age of Authenticity" (Taylor, 2007, p. 514). The word has penetrated deeply into popular culture. Creating the "sensation of authenticity" is a billion-dollar business and one of the fundaments of consumer economics (Gilmore and Pine II, 2007).

In the western world authenticity is often considered as a natural characteristic of human beings by many people. When a person is authentic, he is close to his *true nature*. On the internet you can find an enormous amount of authenticity tests for one's personality. There is a nice paradox in assessing a person's authenticity. The more authentic you want to be, the less your authenticity will be experienced as such by others. Your authenticity tends to withdraw and disappear when you are too conscious of it. Authenticity "disappears when you talk about it" (Weijts, 2016, p. 61). Ankersmit relates a feeling of authenticity to an experience of something being imperfect. What is perfect will never affect us. (Ankersmit, 1993, p. 24).

---

[1] I would like to express my gratitude to Geert-Jan van Bussel, Arnoud Glaudemans and Martijn van Otterlo for their comments on earlier versions.

Knaller has written about *Esthetic Authenticity* and distinguishes *Subject Authenticity, Art Authenticity* and *Referential Authenticity*, during which she has uncovered many paradoxes. (Knaller, 2007, p. 21-24)

Lionel Trilling (1972) has described how the word authenticity got adopted for describing the nature of people:

> ".... we are impelled to use some word which denotes the nature of this being and which accounts for the high value we put upon it. The word we employ for this purpose is 'authenticity'. It is a word of ominous import. As we use it in reference to human existence, its provenance is the museum, where persons expert in such matters test whether objects of art are what they appear to or are claimed to be, and therefore worth the price that is asked for them – or, if this has already been paid, worth the admiration they are being given. That the word has become part of the moral slang of our day points to the peculiar nature of our fallen condition, our anxiety over the credibility of existence and of individual existences. An eighteenth-century aesthetician states our concern succinctly – 'Born originals', Edward Young said, 'how comes it to pass that we die in copies?' "(p. 81).

Authenticity is used by Luciano Floridi as a qualification of an identity of human beings in the infosphere. (Floridi, 2015). His use is rooted in western tradition of assessing authenticity in two ways: how can you trust that someone can be identified in a reliable way, and how can you trust someone acts like his "authentic self", like Trilling describes in the citation above? In Floridi's *4th Revolution*, the word evidently does not lose its importance and continues to have meanings and functions that have been in use for a long time.

Whenever the word authenticity is used as a signifier for the trustworthiness of an object, it is considered a fundamental value in our society. If the authenticity of a document or artefact cannot be proved, it is considered a fake. Lives have been saved by proving the authenticity of documents. Lies have been uncovered by proving that a work of art, or a medieval charter is a fake. History has been rewritten when sources turned out to be unreliable.

To add some complexity: authenticity is a word that is rooted in western cultures. In a lot of non-western languages, the word authenticity did not even exist (Falser, 2012, p. 77). In post-colonial history the word authenticity is often perceived as colonialist and humiliating (Bhatti, 2012, p. 61). The tourist industry is an example to see what the sometimes disastrous implications of "searching for authenticity" may have for whole societies.

Authenticity is a sticky, slippery and self-contradictory word. It is as empty and as meaningful as you want it to be. It is linked to evidence, trustworthiness, justice, esthetic sensation, moral judgements, propaganda, marketing, colonial oppression and to making lots of money.

The purpose of this essay and definitions of key terms

The question that this essay tries to answer is: can this confusing word "authenticity" be an applicable qualification for records in a digital environment? To answer this question, the following sub-questions will be raised:

1. How is the term authenticity used in respect to records?
2. Does the nature of records change in a digital environment?
3. Can authenticity still be used as a qualification even when the nature of records changes in a digital environment?
4. Can we speak of a paradigm shift regarding the term authenticity in a digital environment?

To search for answers, I will introduce a framework for authentication and I will frequently use insights from disciplines like archival theory, art history, information philosophy and ecological philosophy. Since this essay is an attempt to explore new perspectives it inevitably contains elements of speculation.

In this essay the terms *authentic, authenticity, authentication, authenticator, authenticated and informed* will be used as follows:

1. Authentic is defined in The Stanford Encyclopedia of Philosophy as follows: "The term 'authentic' is used either in the strong sense of being "of undisputed origin or authorship", or in a weaker sense of being "faithful to an original" or a "reliable, accurate representation". To say that something is authentic is to say that it is what it professes to be, or what it is reputed to be, in origin or authorship." [2];
2. Authenticity is a qualification granted to an object, a person, a group, an event or another entity. The value may consist of being of "what it purports to be" (Park, 2001, p. 272), or of contributing to the experience of the authenticity of another object, person or a group. The former is the traditional definition as used in diplomatics. In this essay it will be called *Strong Authenticity*. The latter is the definition of what Susanne Knaller has called esthetic authenticity (Knaller, 2007, p. 21-24). It will be called *Weak Authenticity*;
3. Authentication is the process that results in the acceptance or rejection of a claim of authenticity;
4. The *Authenticator* is a person, group machine or other entity that asserts, confirms of rejects a claim of authenticity;
5. The *Authenticated* is a person, group, object, event or other entity for which a claim for authenticity is asserted, confirmed or rejected;
6. The *Informed* is a person, group, machine or other entity that receives the result of authentication.

In the following paragraphs I will explore how the nature of records (the authenticated) changes in a digital environment and how the interaction between authenticator, authenticated and informed may function in this environment. Subsequently the possible interactions between these entities will be described. The essay concludes with an attempt to answer the questions and with observations on change and continuity.

---

[2] https://plato.stanford.edu/entries/authenticity/, accessed 29-09-2017

Records in a digital environment: from object to hyperobject

When we think about authenticity and authentication it is inevitable that questions must be addressed on *what* is to be authenticated. In this essay we focus on digital records. In this paragraph I will describe my position. That position is built on a line of reasoning that is derived from a probably unlikely combination of the work of scholars from different disciplines. At the end of the paragraph I will introduce the idea of digital records being a part of a hyperobject containing digital information.

What is a record?

There is a huge amount of definitions of record, and the definitions have shifted over time. Geoffrey Yeo (2008) has written:

> "(…) it now seems appropriate to characterize records as persistent representations of activities or other occurrents, created by participants or observers of those occurrents or by their proxies; or sets of such representations representing particular occurrents" (p. 136).

Yeo's definition is used here because it does not give any restrictions on the form in which the representation takes place, on who authenticates the record, on who decides to create and manage the record and on whether the record should serve as evidence or should serve other purposes. Furthermore, the definition is inclusive regarding both strong and weak authenticity.

The definition gives a clear framework for understanding records. However, it does not include a statement on what the representation *consists of*. In my opinion the representation consists of information, as understood by Luciano Floridi. The *General Definition of Information* states that an instance of information consists of *well formed, meaningful data, "well formed"* meaning that the data are "rightly put together according to the rules (syntax) that govern the chosen system" (Floridi, 2010, p. 20-21). This definition is very appropriate for records, since it does not impose any limitations on the form of the representation as well. It is applicable for medieval charters, for oral records from indigenous communities and for relational databases.

In my opinion this addition should be made to Yeo's definition of a record as only then it is possible to make general statements on questions like: *when is a record*, and *where is a record?*

When is a record?

When we reason according to the definition of *records* given above, we might imply that records are the result of data processing. The persistent representation that results out of this data processing can take on any form as long as the result is acceptable for the user. This user can be a human being, a machine or a network consisting of connected machines and human beings. In this way it becomes hard to talk about a record in the traditional sense.

Records, being the result of processing, can be compared to installation art. They consist of elements (data) that together express what the *creator* intended to express. For example, take a video installation from the 80-ies of the last century, comprising a VHS-videorecorder, an old TV and some fluorescent tubes. The parts

of this installation will break down relatively quickly, especially when they should function every day during the opening hours of a museum. The TV will break down, the fluorescent tubes will fail and of course the VHS-recorder will come to its end after some years. So, to maintain the authenticity, parts will have to be replaced. And then the curator will have to have luck on his side: maybe the fluorescent tubes are not in stock anymore, VHS recorders have disappeared from the earth and of course the TV is so unique that it should deserve its own place in a technology museum. So, after some years what will be left of the *authenticity* of the installation? It seems that the authenticity of the concept can only be maintained by undermining the authenticity of the original material. Maybe the only good measure is to document and to account for all the preservation actions the curator has undertaken. The curator can never "freeze" the installation in such a way that it will always stay the same. Van Saaze therefore considers authenticity the result of activities concerned with conserving the material and by documenting these activities and to give a framework to assess the quality of these activities. (Van Saaze, 2013, p. 80-83)

Just like installation art, digital records can never be preserved in such a way that they "always stays the same". They would vanish within a few years. An archivist should make sure that the information will be available and accessible for as many years as possible. The data must be migrated and converted. Data must be added. Formats must be substituted. All these processes must be documented. To keep the record authentic, it might even be necessary to violate the authenticity of its components: the data.

A record could therefore be considered as a dynamic result of dataprocessing. Acker has written:

> "I would argue that the project of defining a record must be abandoned in the age of networked record (…) in the digital age the ontological purity and drive for the true nature of the record is over. It cannot be exclusively or definitely located because it is not in a single place, it never ends, and is 'always in a process of becoming' in the continuum" (Acker, 2016, p. 316).

However, Acker's position is debatable: whereas processing and re-processing records can be described as "a process of becoming", the digital or analog record, once created and stored, should stay the same in its persistent representation of activities.

Where is a record?

If records are the results of data processing, is it still possible to speak about records as being objects that can be located? It is possible in the classic situation where records are fixed on information carriers, stored in physical repositories. However, it is obvious that it is very difficult to locate digital born records in this way.

When records are considered the result of data processing, it might be argued that a digital informational object does not exist at all, at least in the way physical objects are defined within the classic view on paper records. Of course, digital data are always stored somewhere. In that sense digital information is as material as analog information. However, the place of storage of all the components that make up for digital information (data and software) can vary endlessly.

Digitally stored records must always be processed to become available. The activity of pushing a button and thereby executing one or more algorithms, is essentially different from going to the repository and taking the file from the shelf (Flusser, 2014, p. 31). *Time-binding systems* will produce persistent representations by processing data and offering available, accessible and authenticated, records.

Another consideration is that with the disappearance of physicality of the object the difference between an original and a copy may have become obsolete. Walter Benjamin wrote about the difference between the mastertape and the copies that were distributed for use in cinemas in "the age of reproduction". The *auratic* quality of the masters, according to Benjamin, made them more authentic than the copies (Benjamin, 2009, p. 233). In a digital world there is not such a thing as an original or a copy anymore. You only have results of the processing of data that constitute the record. The record exists when it is used, and it withdraws once the process of using the record is completed. Its components: the data, will be put together by algorithms again when there is a new need for using the record. Where there is no original, there is no copy either. So, unlike the human beings of Lionel Trilling in the opening paragraph: no digital record will be born original and die a copy.

It seems that we cannot talk of objects on the level of abstraction like we used to in a world where information is fixed on analog carriers. Still, when we assume that an authentication of records needs something that can be called an "object" then we might turn to stratification strategies, as for example Nikolai Hartmann proposed (Störig, 2010, p. 592-593). The idea that an object consists of a series of layers is well-known, for example in network and communication protocols. The various layers in a digital environment, like the bitstream, the coding needed to interpret the bitstream, the content, the form, the internal functions, the internal structure, the format and the means and ways of presentation, constitute the object together. If all these strata are preserved convincingly enough for the authenticator and the informed, then we might speak of the processing of an authentic digital record.

The record in a digital environment, being the result of data processing, is often not located in one place. It is often not managed by one manager, not authenticated by one authenticator. It can take an enormous amount of different shapes. In cloud environments, the data are scattered and moved around every second on a global scale. Records are literally everywhere and nowhere. They appear as results of data processing and they withdraw again once they have been used, and they will be processed again whenever they are needed.

The Records Continuum and the Hyperobject

Thinking in lifecycles is a common way of thinking about records in a wide array of disciplines. Floridi has used the lifecycle model as well (Floridi,2010, p. 4-5). However, in archival theory it has become a commonly accepted view since the 1990-ies that lifecycle concepts do not longer suffice as a basis for analysing the creation, the organisation and the use of records. Continuum thinking, as in the Records Continuum model has become a dominant conceptual framework.

According to Upward, the *Records Continuum Model* has the following starting-points:

> "*1. A concept of 'records' which is inclusive of records of continuing value (= archives), which stresses their uses for transactional, evidentiary and memory purposes, and which unifies approaches to archiving/recordkeeping whether records are kept for a split second or a millennium;*
> *2. A focus on records as logical rather than physical entities, regardless of whether they are in paper or electronic form;*
> *3. Institutionalization of the recordkeeping profession's role requires a particular emphasis on the need to integrate recordkeeping into business and societal processes and purposes.*" (Upward, 1996, p. 4-5)

The Records Continuum Model implies that we might fully understand, analyse and control the context and content of records. To understand issues of authenticity of records in a digital environment, we might rethink this claim of records. A helpful, new perspective is offered in *ecological philosophy*, especially by Timothy Morton.

Morton, who is a member of the object-oriented philosophy movement, argues that it is philosophically sound to think in terms of "hyperobjects". These are objects that exist beyond the possibility of humans to perceive. They are too massively distributed in spacetime. According to Morton it is necessary to think in this kind of concepts to understand major ecological developments like climate change. The concept has not been applied to ecological issues only. For example, it has been used by the Dutch philosopher René ten Bos (2015) to gain a deeper understanding of a very different subject: bureaucracy. In this spirit I will try to apply Morton's concept of hyperobjects to records.

According to Morton, hyperobjects are so vast that human beings are not capable to oversee and comprehend them. It is impossible to get a grip on them. Human beings are immersed in them. Parts of the hyperobject may be revealed but human beings are not capable to grasp them in full: "It is as if we were inside a gigantic octopus" (Ten Bos, 2015, p. 36-37).

Records can be considered nowadays a part of a hyperobject consisting of digital data and information. Whole societies are immersed in digital data and have become dependent from them. Nobody will be able to oversee all the data that are created and managed by for example the Internet of Things and peer-to-peer technologies. Every human being, every machine and even every network will only get partial and temporal glimpses of this digital information.

According to Morton (2012) hyperobjects share five properties. I will try to apply them to a digital environment:

1. The first one is *viscosity*: hyperobjects stick to us, no matter how hard we try to resist. Nobody will be able to avoid the influence of digital data and records. It is mainly a one-way direction: the hyperobject influences your life, and your influence is very restricted. There is a massive number of examples on how digital data, including digital records, are used and re-used and thus involuntarily influencing everyone's.

2. A second feature of hyperobjects is that they are *non-local*. They cannot be pointed to a specific time and place. Every archivist will recognise this as a property of digital records. The data that are used to process records can be stored anywhere. They can be used anytime and in any place. This feature already exists at the most basic level of an information object. Cloud storage that is nowadays implemented by for example Google, Amazon and Microsoft, chops the bitstream of information objects in different pieces and dynamically stores and shifts each piece into different datacentres all around the globe.

3. A third feature of hyperobjects is their different *temporality*. They are so vastly distributed in time that they "force us to drop time as a neutral container". Hyperobjects even "emit" time, just like planets do. The notion that records transcend the timescale of a human life is basic to every archivist and to every user of historical information. However, this feature implies a shift from recent concepts like the Records Continuum. Hyperobjects are no continuums. They just exist in bigger timescales and in bigger life cycles than the ones we are familiar with.

4. Furthermore, hyperobjects are *phased*. That means they possess high-dimensional phase spaces that makes it practically impossible for a human to assess them accurately. When you think of digital information you might think of the problems of the whereabouts of the data. In a cloud environment, as mentioned before, they can be anywhere. There are enormous amounts of data necessary to assess not only the whereabouts, but also the availability, accessibility and authenticity. The content and the whereabouts of these data may vary every nanosecond. It seems that this characteristic can also be applied to paper records. Paper records disappear and re-appear over time as well. For example, during the renovation of a house in the German town of Erfurt in 2016, records were discovered that were created during the sterilisation campaigns of the Nazi-regime.[3] These records were never destroyed. They had been available all the time. However, they were not accessible. They were therefore unknown to the world. Moreover, archeologists use the term "soil archive" in a comparable context.

5. The last property that Morton attributes to a hyperobject, is that of *interobjectivity*. They are formed by relations between objects. They consist of these objects, but they are not reduced to one. The information that results from these relations make the hyperobject visible. What comes to the mind of the archivist here is the keyword *context*.

Hyperobjects transcend human possibilities to oversee them in full. Yet they have an enormous impact on our lives. We cannot withdraw ourselves from their influence. The resemblance with digital records is striking. They are by their very nature contextual. If they are managed well they transcend our own timescale. Digital data are so vast that it is hardly possible to get a grip on their whereabouts. They are not bound to specific locations. And their use is time- and location independent. And records are sticky: they will influence you in any time and in any place. Digital records also have the tendency to withdraw: once they have been made manifest, they tend to hide again (in the cloud for example), like an octopus.

A digital hyperobject is sticky and slippery. And digital records might be part of this hyperobject. To determine the stickiness of this octopus, we might use a familiar word: authenticity. However: how to authenticate a data processing hyperobject?

Authenticating records of a hyperobject: towards self-authentication?

One might expect that the complex meaning of the word authenticity can be reduced once it is related to records. However, this is not the case. The reduction of our scope leads to something that resembles a Tardis. It looks small from the outside but when you open it, a vast world, almost without spatial boundaries, opens. In respect to records, the word authenticity remains sticky, slippery and does not lead to simple and clear notions.

For example: when an archival institute digitises analog information in a correct and legitimate way, it can be criticized for diminishing its authenticity: the loss of the original material.[4] A collection of records documenting the history of a city, can be authentic in the strong meaning of the word. Still it can be criticized as being non-authentic in the weak meaning of the word if the information does not represent enough parts of the history and the present society. Another example: if a collection of digital records is preserved in a sound, reliable way then strong authenticity is debatable. The "original" will have to be transformed almost continually to keep the information available and accessible.

All these examples involve paradoxes. One way to try to overcome these paradoxes is to look at the way the verdict of "authentic" is reached and who reaches that verdict. Authentication is a validation process leading to the conclusion that something or someone is genuine, true and/or reliable (or not). It leads to the assumption whether we are talking about a "Real Thing" or a "Fake". To make sense of the enormous possibilities and contexts of authenticity the method described below might be of use. It is an analysis of the process leading to the verdict.

Three entities are involved in the process: the authenticator, the authenticated and the informed. The *authenticated* might be a human being, a group or organisation, a physical object or a digital object. Later I will get back to the issue whether digital information is an object at all, since it is a relevant issue in the process of authentication. The *authenticator* and the *informed* are both *agents*, and can be human beings or machines (Floridi, 2010, p. 103). The authenticator, the authenticated and the informed might even blend into one agent.

External and internal authenticators

Authentication used to be a process that involved an *external* authority. This authenticator could be of divine origine. When the authenticator is a human being, it often is a notary, an archivist, a philologist or a leading figure in the community.

---

[3] https://www.welt.de/kultur/history/article157586184/Bauarbeiter-entdecken-eingemauerte-NS-Akten.html, accessed 29-9-2017

[4] For example: digitizing records and disposing the originals, was vehemently disputed in a Dutch newspaper by Arnold Heertje, http://www.parool.nl/opinie/-weg-met-het-intellectuele-tekort-van-hoog-tot-laag~a4347115/, accessed 29-9-2017

The origin of the words *authentic*[5] and *genuine*[6] point to the idea that authentication had to do with oppressive violence. It was the authenticator who had supreme power to tell what is true or not. The truth: whether something was trustworthy and genuine, was to be found outside the authenticated and was to be imposed upon the informed. However, even in this very traditional and strict interpretation the Informed still played a role: authenticity only exists when the informed is convinced.

Since the eighteenth century the true source of authenticity has increasingly been *inside* the authenticated. There was a true and pure inner self that was to be sought and found. This western romantic idealism tended to become more important when the power of (authoritarian) religion diminished and when traditional authorities became more and more controversial. The truth could not to be imposed upon by some authority. The authenticator used to be *exoteric* and became more and more *esoteric* (Sloterdijk, 2016, p. 106-107). Authenticity was to be found within yourself and only by yourself.

In this last sense authenticity is something that should be attained, never to be completely reached. It withdraws itself when it becomes too obvious, as described above. The authenticator becomes more and more identical with the authenticated, and with the informed. Moral philosophers like Charles Taylor (2007) have built their case on this: there is no external, divine authority anymore, we should look for authentication within ourselves. We are our own authenticators. And we are all in Plato's cave.

Strong and weak authenticity

The best-known use of the word authenticity in relation to records is: there is an object, for example a medieval manuscript, and an authority will determine whether it is genuine. In this case authenticity is an essential qualification for the trustworthiness of the object. It identifies its nature. Following the definition in the Stanford Encyclopedia of Philosophy cited at the beginning of this essay, the term *strong authenticity* will be used for this type.

Authentication might also be of referential nature. In this case the authentication process is directed towards the question if an object or person is a true, trustworthy *representation* of something, or someone, else. Does a collection of records give a trustworthy and representative picture of what happened in a city, a country, in the life of a person or family? In this essay the term *weak authenticity* is used to describe this type.

Weak authenticity is closely connected to questions of identity: the informed might be very satisfied when referential authenticity confirms his or her own opinions and worldviews. Weak authenticity where the informed and the authenticator are one and the same, is the billion-dollar business Gilmore and Pine (2007) have described. The search is a powerful catalyst of consumer economics: we want to surround ourselves with things that reflect and confirm our identity. "This is so not us …" as a rather funny TV car-commercial puts it.[7]

When we combine the types of authenticity described above, we are provided with four possibilities for authentication:

1. *External* judgement of *strong* authenticity;
2. *External* judgement of *weak* authenticity;
3. *Internal* judgement of *weak* authenticity;
4. *Internal* judgement of *strong* authenticity.

These possibilities will be briefly explored below.

External judgement of strong authenticity

The first possibility is the best known to archivists. It is the root of archival science: *diplomatics*. Authentication is the validation of records with the aim to determine if they can be trusted: "When a record is what it purports to be, the record is genuine" (Park, 2001, p. 272). Whenever the validation is positive the record is a reliable source. If so, the record can be used for evidential purposes (Duranti, 1995, p. 6). Strong authenticity of a record must be proved by a test of specific conditions. The test should always give the same results under the same conditions.

The study of language in written sources at least goes back to Antiquity and to the Islamic Golden Age. In Europe philology became prominent during the Renaissance by scholars like Lorenzo Valla and Erasmus. Jean Mabillon established diplomatics as a research method in the sixteenth century. Diplomatics became a positivist science in the nineteenth century. The scientific method attempted to uncover general principles under an empirical reality through scientific experiments. As MacNeil (2016) wrote: "The diplomatic analysis of the elements of a record is a process of abstraction and systemization, the aim of which is to identify the essential attributes of a record and make them transportable to different historical and documentary contexts" (p. 734).

The way in which authentication takes place has varied a lot, depending on who was the authenticator. It might be a priestess in Delphi, it might be the Pope, a notary, a judge or an archivist. Nowadays the authenticator might be a machine, or a huge network of machines. Authentication using blockchain technology is the reality now. Authentication of records in the Internet of Things and peer-to-peer technologies will become reality without a doubt. *Authentication by Design* will become a necessary part of designing information systems. The necessary precondition will be that enough (meta-)data are added to the record during creation and during preservation. This includes data about the kind of algorithms being used, about rules for preservation, about parameters for authentication and about restrictions on access.

Based on the results of the *InterPares*-project, MacNeil (2016) has claimed that diplomatics in a strict positivist sense can no longer hold the positivist claim that it proves strong authenticity of records. Reality has become too complicated, especially

---

[5] See http://labyrinth.rienkjonker.nl/content/herkomst-authenticiteit, accessed 03-06-2017, citing *Beknopt Grieks-Nederlands woordenboek*, (1969) bewerkt door dr. W. den Boer, Wolters Noordhoff NV, Groningen, p. 134

[6] See http://www.perseus.tufts.edu/hopper/morph?l=genuinus&la=la&can=genuinus0#lexicon, accessed 03-06-2017

[7] See https://youtu.be/CBPguh2JfBw, accessed 29-9-2017

for records in a digital environment. Instead of a positivist approach, MacNeil (2016) proposed that diplomatics as an analytical method should be based on a more interpretivist approach. This also means a shift in the way the word authenticity is used. It does not reflect an "objective truth" but "an understanding that reflects fairly the various perspectives of participants in that setting" (p. 755).

Even though innovative processes of authentication are implemented, and whatever positivist or interpretivist background is used, the idea behind them remains that the authenticator is an external informational agent that validates the record and makes the judgement of authenticity.

External judgement of weak authenticity

The second interpretation of authenticity as a property of records is of referential nature. Weak authenticity is concerned with the question whether records are a sound representation of the history of for example a person, a city, an organisation or a country. This interpretation is often linked with political issues. The authenticator controls the creation, the appraisal, the disposal and the distribution of records. The authenticator therefore controls the sources, the evidence, upon which historical research is based.

In the past decades weak authenticity, and the power of the authenticator, have become an important subject of research. As McKemmish and Gilliland (2016) have written: "The greater diversity and the expanding research front reflect in part the impact of the so-called 'archival turn' first evident in postmodern and postcolonial discourses in disciplines like anthropology, literature and history. It has encouraged researchers in archival science to contemplate the societal implications and effects of archives and recordkeeping. Critical theory (…) provides a framework for theorizing about both the role of the Archive in social conditions and forces such as colonialism, oppression, marginalization and abuse of human rights, and the part that it might play in postcolonial, post-trauma and post-conflict societies. Increasingly the Archive is being explored as a contested, political space, a societal concept associated with the promotion of asymmetrical power, grand narratives, nationalism, surveillance, and the omission, diminution or silencing of alternate narratives (…) Interdisciplinary areas such as race and ethnicity, gender and ion, and Indigenous and studies, are also addressing the role of the Archive (…)". (p. 86)

An example of an authenticator is the archival institute that selects and manages records of a city. Every institute will have its appraisal policy, and a lot of them will try to collect data and records from every part of society. In this case the archivist is the gatekeeper, classifying records as fit for purpose.
The issues can be highly political as is clear from the citation above. The authenticator can deliberately contribute to inequality, for example when property rights of indigenous communities in Africa and in Northern America are not recorded in a way that the ruling class view as reliable and thus authentic (Faulkhead and Thorpe, 2016). Records can also contribute to inequality when for example abuses of power by for example military invaders or domestic oppressors are not recorded and captured.[8]

---

[8] See for example the work of witness.org, accessed 29-09-2017

Another, less politically loaded, example is the use of records for city marketing. The historical museum of the city where I live, creates a very positive view of the history of the city (with keywords like tolerant, entrepreneurial and creative) and shows records and artefacts to illustrate this. For this purpose, it is quite inconvenient that the museum contains a lot of records and artefacts about events that the city cannot be proud of, like the city's contribution to slave trade until the nineteenth century. This is dealt with in one short sentence at the end of the statement about the "DNA of Amsterdam". Records are put in the context of city branding,[9] or to put it less mildly: in the context of outright propaganda and distortion of historical truth for financial benefits.

However, the notion that for this kind of authentication you need sources that have proved to be reliable and genuine, remains an important matter. Weak authenticity needs records with strong authenticity.

Internal judgement of weak authenticity

This is a postmodernist, constructivist subject. The link with strong authenticity is practically absent. This kind of authentication only contributes to what the authenticator wants to see confirmed in authenticating himself. If it looks or sounds like a record, and if contributes to the identity the authenticator wishes for himself, it is OK. If the authenticator gets a feeling of his own authenticity by using the record, it is fine. Authenticity, when authorized from the inside, does not need authentication from the outside. Anything that strengthens the idea that one comes closer to himself, will do.

This is the realm of consumerist behaviour as described by Gilmore and Pine (2007). It is a postmodernist realm where values like truth, reliability and proof are always "subjective" and are always dependant on "metanarratives". This is also the realm of contemporary discussions about fake news, filter bubbles and possibilities of manipulating digital texts, photos, videos and animations. It is therefore a highly relevant subject that needs further exploration.

Internal judgement of strong authenticity

The three ways of authentication that are described above radically differ from each other. The first is traditionally linked to diplomatics. The second poses questions to the function of records in society. The third is concerned with weak authenticity when strong authenticity is irrelevant. All these types of authentication have been researched.

The fourth possibility is that of the record claiming, accepting or rejecting its own strong authenticity. These records are part of an environment that does not allow for an outside authenticator, because they are part of a hyperobject that is too vast to control by any individual or group, machine or network.

Digital records might beyond the control we were used to in the analog environment. Laws, regulations and standards will only partially help. Getting a grip on a hyperobject is a self-contradictory goal. Therefore, the only realistic authentication process of digital records might be internal. The octopus should be

---

[9] See http://hart.amsterdammuseum.nl/nl/page/35365/amsterdam-dna-tijdlijn, accessed 03-06-2017

able to authenticate itself. Nobody else can do it. The authenticator must be inside the authenticated. Human beings should help the authenticator for doing its job right. We should enable the hyperobject to authenticate itself.

This is one step further than *Authentication by Design*. Authentication of a hyperobject presupposes artificial intelligence that will enable records to learn to authenticate themselves. There is no outside interference in the authentication process anymore.

Concluding remarks

The question that this essay tries to answer is: can this confusing word, authenticity, be an applicable qualification for records in a digital environment? To answer this question, the following sub-questions were raised:
1. How is the term authenticity used in respect to records?
2. Does the nature of records change in a digital environment?
3. Can authenticity still be used as a qualification even when the nature of records changes in a digital environment?
4. Can we speak of a paradigm shift regarding the term authenticity in a digital environment?

At this point, the answers to these questions are in my opinion:
1. It is used in a strong meaning, being a qualification of the nature of a record, and in a weak meaning, being a qualification of the way in which the record confirms the identity of another entity, for example a person, a group, an object or an event;
2. Yes, they have become the results of data processing;
3. Yes, even though authentication of digital records (when considered part of a hyperobject) is a process that might need new technologies;
4. No, see below.

In my view, *authenticity* therefore remains an applicable qualifier for records. However brief and sometimes speculative this essay is, my conclusion is that, even if we see digital records as part of a hyperobject on a totally new scale, both strong and weak authenticity continue to be of vital importance. Authenticity remains a very sticky word. In my view it will continue to play a crucial role, despite all complications, paradoxes and obscurities that some with it.

As to the last sub-question the answer is that there is no paradigm shift. There is no incommensurability between using the word in the context of analog records and in the context of digital records. It seems that the enormous *acceleration* (or growth) of records made possible by the changed nature of records in a digital environment, is the level on which a revolution is taking place (Sloterdijk, 2016, p. 15-20). It is like we are witnessing the emergence of an uncontrollable Open Archival Information System (OAIS) running on steroids. That would mean that the incommensurability of words does not take place on a fundamental conceptual level, but on logical, methodological and instrumental levels.

I would like to conclude this essay with some closing remarks.

- Authentication is a highly ethical process. The verdict of authenticity is based on values of reliability and trustworthiness, either of the record itself or of the reference made to the record.
  - *Technological issues* will be predominant in authentication processes concerning the nature of records. How far can technology go to be able to *Authenticate by Design* and which parts of technology should be affected by it? Will it affect the algorithms, the integrity of the data and the bitstreams? And will it be possible to create artificial intelligence for records, so they can learn how to authenticate themselves?
  - *Issues on power and awareness* will be predominant in all authentication processes. Who owns the future in terms of strong authenticity of records? And who owns the future concerning the distribution of records seen from the perspective of weak authenticity?

- To explore the ethical challenges Floridi's ethical RPT model can be useful (Floridi, 2010, p. 104). This model gives the opportunity to define ethical actions of informational agents towards digital records. It makes a distinction of three types of actions: using information as *Resource, Product* and *Target*. Especially in producing (or: creating) and targeting (or: management) of digital records there is lot of research to be done to enable the authenticator.

- Both the hyperobject and its qualifications of strong and weak authenticity have the tendency to withdraw after becoming manifest. They play a game of assessing, appearing and disappearing. The game is highly relevant in present day discussions about filter bubbles, fake news and accessibility to evidence. Records are not about whether the information is true. However, the claim of truth can only be made by referring to sources (including records), that have a strong authenticity.

- In this essay I have avoided to use the word "archive", to concentrate on its constituent parts: records. The reason is that there are many definitions of an "archive" for example: the archive as the physical place where records are kept, the archive as the recordkeeping institute, the archive as the collection of records that constitute the representation of an organizations or the archive as the collection of records that represent the memory of a city of a country. Each definition would require its own considerations regarding authenticity in a digital environment and regarding high level concepts as hyperobjects (Morton, 2013) and infosphere (Floridi, 2010).

- The nature of records has changed dramatically on different levels of abstraction: being transformed from locked-up, scarce information, that has survived natural and man-made destruction, sticked to material and put away in specific locations, into innumerable objects of information that fit well into the five properties Morton describes for hyperobjects. The image of digital records as being part of a hyperobject consisting of digital information fits well into Floridi's concept of the fourth revolution as a "blow to anthropocentrism" (Morton, 2013, p. 36).

We are immersed in digital information. Every bit of this hyperobject affects us. Helping to maintain the strong authenticity of records that are part of it, should be a part of our ethical obligations. Nobody can express the importance of this in a better way than Jorge Luis Borges (2015) did in his poem The Plot, that ends with the lines: "Of all those lost things there is not one that does not throw a long shadow and that does not determine what you do today or what you will do tomorrow".[10]

Literature

Acker, Amelia (2016). When is a record; a research framework for locating electronic records in infrastructure. In A.J. Gilliland, S. McKemmish, A.J. Lau (eds.), *Research in the Archival Multiverse* (pp. 288-323). Clayton, VIC: Monash University Publishing. http://dx.doi.org/10.26530/OAPEN_628143

Ankersmit, Frans (1993). *De Historische Ervaring*. Groningen: Historische Uitgeverij.

Benjamin, Walter (2009). The work of art in the age of mechanical reproduction. In *One-Way Street and Other Writings* (transl. J.A. Underwood, pp. 228-259). London: Penguin Classics.

Bhatti, Anil (2012). Die Abwendung von der Authentizität, postkoloniale Sichtweisen in neueren Romanen. In Michael Rössner, Heidemarie Uhl (eds.), Renaissance der Authentizität? Über die neue Sehnsucht nach dem Ursprünglichen (pp. 43-62). Bielefeld: Transcript.

Borges, Jorge Luis (2015). *Alle Gedichten* (transl. Into Dutch Barber van der Pol en Maarten Steenmeijer). Amsterdam: De Bezige Bij.

Bos, René ten (2015). *Bureaucratie is een inktvis*. Amsterdam: Boom uitgevers.

Duranti, L. (1995). Reliability and Authenticity: The concepts and their implications. *Archivaria, 39*, 5-10.

Falser, Michael S. (2012). Von der Charta von Venedig 1964 zum NARA Document on Authenticity 1994. In Michael Rössner, Heidemarie Uhl (eds.), *Renaissance der Authentizität? Über die neue Sehnsucht nach dem Ursprünglichen* (pp. 63-88). Bielefeld: Transcript.

Faulkhead, Shannon, and Kirsten Thorpe (2016). Dedication: Archives and indigenous communities. In A.J. Gilliland, S. McKemmish, A.J. Lau (eds.), *Research in the Archival Multiverse* (pp. 2-15). Clayton, VIC: Monash University Publishing. http://dx.doi.org/10.26530/OAPEN_628143

Floridi, L. (2010). *Information, A Very Short Introduction*. Oxford: Oxford University Press.

Floridi, L. (2015). A Proxy Culture. *Journal for Philosophy and Technology, 28*(4), 487-490.

Flusser, Vilem (2014). *In het universum van de technische beelden* (transl. Into Dutch M. Geerards). Utrecht: Uitgeverij IJzer.

Gilmore, James H., & B. Joseph Pine II (2007). *Authenticity. What Consumers really want*. Boston: Harvard Business Review Press.

Knaller, Susanne (2007). *Ein Wort aus der Fremde, Geschichte und Theorie des Begriffs Authentizität*. Heidelberg: Winter.

McKemmish, Sue, & Anne Gilliland (2013). Archival and Recordkeeping Research; Past, present and future. In Graeme Johanson, Kirsty Williamson (eds.), *Methods: Information, Systems and Contexts* (pp. 79 -112). Prahran, VIC: Tilde Publishing and Distribution.

MacNeil, Heather (2016). Contemporary Archival Diplomatics as a Method of Inquiry. In A.J. Gilliland, S. McKemmish, A.J. Lau (eds.), *Research in the Archival Multiverse* (pp. 724-760). Clayton, VIC: Monash University Publishing. http://dx.doi.org/10.26530/OAPEN_628143

---

[10] "Van die verloren dingen is er geen dat niet op dit moment een lange schaduw werpt en niet bepaalt wat je vandaag doet of wat je morgen zult doen." (Jorge Luis Borges-De Plot in: Alle Gedichten, pag, 578. De Bezige Bij, 2015). English translation by the author of this essay.

Morton, Timothy (2012). *What do Hyperobjects say?* Retrieved from
http://ecologywithoutnature.blogspot.nl/2012/12/what-does-hyperobjects-say.html

Morton, Timothy (2013). *Hyperobjects, Philosophy and Ecology after the End of the World*.
Minneapolis: University of Minnesota Press.

Park, Eun (2001). Understanding "Authenticity" in Records and Information management:
Analysing Practioner Constructs. *The American Archivist, 63*, 72-91

Saaze, Vivian van (2013). *Installation Art and The Museum, presentation and conservation of
changing artworks*. Amsterdam: Amsterdam University Press. Retrieved from
www.oapen.org/download?type=document&docid=449202

Sloterdijk, Peter (2016). *Was geschah im 20. Jahrhundert?* Berlin: Suhrkamp.

Störig, Hans Joachim (2010). *Geschiedenis van de filosofie* (transl. into Dutch P. Brommer).
Houten: Spectrum.

Taylor, Charles (2007). *A Secular Age*. Cambridge, MA: Belknap Press of Harvard University
Press.

Trilling, Lionel (1972). *Sincerity and Authenticity*. Cambridge, MA: Harvard University Press.

Upward, Frank (1996). Structuring the Records Continuum - Part One: Postcustodial
principles and properties. *Archives & Manuscripts, 24*(2), 268-285.

Weijts, Christiaan (2016). *Groene Amsterdammer, 34*.

Yeo, G. (2008). Concepts of Record(2), Prototypes and boundary objects. *The American
Archivist, 71*, 118-143.

266

MARTIJN VAN OTTERLO

From Intended Archivists to Intentional Algivists.
Ethical codes for humans and machines in the archives[*]

"If the digital future is to be our home, then it is we who must make it so"
Shoshana Zuboff

Overview

Starting from the prediction that someday algorithmic archivists – or as I like to call them: algivists – will work the archives, I describe how ethical thinking in traditional archives could be employed to teach algivists moral values. The essay starts by describing the digitalization of society and archives and how so-called codes of ethics have evolved to define the moral values of archivists, characterized as the intended archivist. I then turn to ethical thinking about algorithms, how different types of algorithms induce entirely new classes of ethical challenges, and argue that a good way to endow algivists with ethical behavior is to employ the same kind of technology, algorithms, to encode ethical values directly in their core beliefs as a bias. This results in the intentional archivist, or the algivist. In this essay I develop a vision on the future of the algorithmic archivist and an idea to obtain algorithms in archives that obey our human ethical values.

(1) The Coming Archivist Singularity[1]

[Some place, some time in the future] It took ages to get permission, but yesterday evening I finally got THE mail. I consider myself lucky, since I really needed access to the archives to finish my article. Other people would ask why an assistant professor in technology ethics would like to see those old-fashioned paper documents about the introduction of Mindbook, the company that grew out of the long-gone Facebook Corporation. Since, their summaries are already on Archipedia. Who is interested in paper documents anymore? Well I am. I never felt comfortable with all this digital… eh stuff… anyway. People are physical, and they like physical things. Well… at least that's my opinion. And besides… I don't trust Archipedia; they have appeared in so many algorithmic trials for information manipulation, but they always use their right-to-silence and nobody is able to crack their summarization code. I need to take a look myself. I enter the red building next to the rocket station and turn right after getting

---

[1] All links appearing in footnotes have been last accessed October 2017.

through the bio-scanner. Paul, a robot from the ALGIVIST-5000 series is waiting at the desk. His emotional module can use an update, I catch myself thinking. I only get a nod and a metallic "Hello, how can I help you?", so much unlike the newer models that can really lighten your day with their big smiles and warm voices. I answer the way I am supposed to do, with a clear question and context: "Hello Paul, I'd like to see all documents containing discussions on the use of advanced mind models, especially whole brain simulations, of Facebook users prior to the formation of Mindbook. I also would like to look at pictures and footage of the meetings that include people from the legal department, and can you please provide me with additional CV information of these people? Thank you." Paul knew from prior contact that I would be coming to the archive myself; otherwise he would have downloaded the interpreted documents, or DOC-INTERPRETs as they call them here, to my communicator. Now he only sends the requested CVs and projects an interactive map of the archive a floor below which will guide me to the right boxes. Since Paul scans and stores all items (including photos and a shallow semantic analysis of texts), and organizes them in the physical space, he knows where I have to go. At least, that is what I have to believe since there is no way of knowing what is in the complete archive. While going downstairs, I sense excitement from my side on how optimized and effective my routing past all the right boxes, 16 in total, is. Five more boxes are off limits for me though. It turns out another researcher has a similar research question in parallel, and his (or her?) combined scientific h-index and social media coverage is so much higher than mine. Also, according to an analysis of the planned social activities in our agendas, and our biophysical energy levels in combination with the predicted moist weather in the next weeks, Paul estimates that I will not put enough hours in my analysis of the documents and my writing anyway. Sure… I need to stop eating snacks and boost my metabolism… but come on… who does Paul think he is? My doctor? According to Paul the overall estimated impact of the other researcher publishing the material alone is higher when I do not interfere. I have no other option than to accept, but I don't think it's fair. Archival robots such as Paul are built to optimize their impact since they too get ranked. Of course, everyone gets ranked, and so are archival robots. Paul needs to optimize the use and costs of the archive while at the same time striking a balance between preventing possible negative impact on the donor organization Mindbook, and stimulating positive impact from researchers and journalists publishing the right kind of information, again according to Mindbook. Oh well… the rest of the day I look at the documents, trying to find what I am looking for. The surveillance-sensors watch my every move while interacting with the documents, which helps them to further optimize the archive, so they say. Well… they sure also use them for the projected advertisements that are appearing on the electronic walls for me. Hey… yes indeed… I do need a snack… my hands are trembling…. How did they know? Oh… never mind.

This scenario may sound like science fiction today, but could be happening in the near future. The *algorithmic* archivist Paul, or **algivist** as I will call it, will be a natural outcome of the *digital age* we are only just starting. It is not a matter *if* all this will happen, but *when*. I define the coming **archivist singularity** as the moment when all core archivist's activities will be replaced by an algivist. Usually *singularity* amounts to general technology (Shanahan, 2015) but here I focus more specifically on the archivist profession. Just like in autonomous cars, we can talk about various *levels*[2] of autonomous algivists: some will only maintain digital archives, some will have a robot body (for physical collections), and some may only

function as an assistant of a human archivist. All will, however, be responsible for *selecting, ordering*, and *documenting* archival documents. Introducing algorithms into our lives, and letting them take over jobs that were exclusively done by humans will cause profound changes in society and requires considerable thought on how to do that in a "good" way. A central question in this essay will be about how we can ensure that algivists will uphold the same moral standards and behavior as their human counterparts who have been our human gatekeepers to societally important information for so long.

Worries about the general singularity, when computers will outsmart "us" in every way possible and may spin out of control relative to our human interests, trigger *existential fears*. It reminds of concerns when another technology was in its initial phases: *nuclear technology*. Albert Einstein warned President Roosevelt in 1939 in a letter[3] for the consequences if *some other nation* (Germany) would obtain the technology for powerful bombs and suggested to start a nuclear program in the United States. The current explosion in *digital technology* and *algorithms* may very well trigger a similar arms race. But before worrying about superintelligence, we should study the many ethical challenges of not-yet-fully-superintelligent algorithms, such as our algivist Paul.

Ethical issues with algorithms arise on a daily basis. For example, Google's search algorithm tagged[4] (photos of) black people as "gorillas", showing either a bias in data or learning procedures, or errors in the application of the tagging algorithm. Autonomously driving cars constantly make mistakes[5] or are not yet fully capable of driving in our complex, physical world. A related case is when algorithms are *deliberately* used for the wrong purposes, such as the Dieselgate[6] case which dealt with cheating software to fool emission tests. Another example with ethical dimensions is Facebook's idea to *predict*[7] potential suicides to proactively aid people, which is in the same direction as Google's recent efforts on *depression detection*.[8] Whether some of such issues may be against current or coming laws, in most cases we can say they are at least *creepy* (Tene and Polonetsky, 2014) since people will be targeted by Google's and Facebook's algorithms as *depressed* or *suicidal*: what consequences will that have? Another creepy example is the Cayla[9] doll which can communicate with children, send their data (voice, things said, possibly video capture) to the manufacturers' servers, and in addition, it can say anything to a child through a microphone. Apart from possible hacks, such "connected" dolls are creepy because they invade (just like smart-TVs and cell phones) the privacy of intimate family life, without doing anything illegal.

Other recent ethical challenges have to do with the typical *gatekeeping role of algorithms* employed by search engines and the like: *fake news*, Pariser's (2011) *filter bubbles* (where algorithms reinforce people's biases), and *censorship*. As an example, Facebook's policy to allow or disallow particular content, essentially implementing a form of censorship[10], raises many ethical issues given their 2 billion user-base.

---

[2] https://en.wikipedia.org/wiki/Autonomous_car#Levels_of_driving_automation

[3] https://en.wikipedia.org/wiki/Einstein%E2%80%93Szil%C3%A1rd_letter

[4] https://www.theverge.com/2015/7/1/8880363/google-apologizes-photos-app-tags-two-black-people-gorillas

[5] https://phys.org/news/2016-09-dutch-police-probe-fatal-tesla.html

[6] https://www.cleanenergywire.org/factsheets/dieselgate-timeline-germanys-car-emissions-fraud-scandal

[7] https://www.wired.com/2017/03/artificial-intelligence-learning-predict-prevent-suicide/

[8] https://www.theguardian.com/commentisfree/2017/aug/25/google-clinical-depression-privacy

[9] https://www.nytimes.com/2017/02/17/technology/cayla-talking-doll-hackers.html?mcubz=1

[10] http://fortune.com/2017/05/22/facebook-censorship-transparency/

Recently some of it has been disclosed[11] but generally it is unclear who decides upon them. Facebook is also active in detecting utterances related to terrorism[12], Google aims to tackle fake news by classifying[13] news sources and marking them, effectively implementing a "soft" version of censorship, and Twitter targets[14] "hate-speech", thereby implementing language (and possibly thought) monitoring on the fly. Big technology companies are starting to recognize the ethical[15] issues, even causing Google to revive Wiener's[16] idea of an emergency button[17] to turn off autonomous systems. Ethical concerns about algorithms, or more generally *artificial intelligence (AI)* (Nilsson, 2010), are still relatively new and come from many directions. Open expressions of concerns by Stephen Hawking, Elon Musk and Bill Gates warn[18] for the unforeseen consequences of widespread use of AI. A letter[19] of concern with "research priorities for robust and beneficial AI" was quickly signed by more than 8000 researchers and practitioners. Individual top AI researchers speak out, such as Tom Dietterich[20]. Big tech companies such as Google, Amazon, IBM and Microsoft announced that they are forming an alliance[21] which "aims to set societal and ethical best practice for AI research". Various academic initiatives[22] arise around the broad topic of "societal implications of algorithms" and the scientific literature on the topic is growing quickly (Mittelstadt et al., 2016). Various authors try to explain the complex interactions between algorithmic technology and society. Van Otterlo (2014a) links *behaviorist psychology* to the way technology now has the means to implement *behavioral conditioning* on a large scale. Zuboff (2015) introduces the "Big Other" as a metaphor to point to the combined logic of capitalism, surveillance and digital technologies such as AI. Morozov[23] sees similar patterns of information capitalism undermining our human democracy. All these analyses go beyond relatively simpler, more isolated, issues such as privacy and data protection, and see the potential influence of algorithms on society as a whole, with profound implications for democracy and free will.

In this essay I explore ethical implications of algorithms in archives, with consequences for *access*. One of my goals is to introduce recent developments in the ethical study of artificial intelligence algorithms to the reader and survey important issues. One argument I develop in this essay is that since "we", as humans are creating these future algivists, we should study their ethical implications before, during and after creation. However, I also argue that maybe it is better to try to *create them in such a way that we can ensure that they will behave according to our own moral values*. How to construct this *ethical algivist*, and how does this fit into more general, scientific developments?

(2) The Digitalization and Algorithmization of Society and Archives

One of the hype terms of this decennium[24] is *big data*. Everywhere around us everything is turned into *digital data* which is thought to be good for health, the economy, the advancement of knowledge, and so on (Mayer-Schönberger, 2013). The promise is that data will allow us to understand, predict and optimize any domain (van Otterlo and Feldberg, 2016). For example, patient data allows us to build *statistical models* to predict diseases, and to *experiment* with novel treatments based on the insights of data, to cure more diseases. Another promise of big data is that it allows one to throw[25] away typical "hypothesis-driven" science, which works *top-down*, and to adopt a more *bottom-up* strategy, which starts with the data and tries to find patterns. Big data is not entirely new: *big data "avant-la-lettre"* can for

example be found in the *Cybersyn project* in Chile in the seventies which was aimed at controlling the economy of a complete country (Medina, 2015), something which sounds like modern "smart city"[26] endeavours. Data has always[27] been gathered and analysed but the scale of today is new. Modern data-driven technology induces a new[28] machine age, or an industrial revolution (see also Floridi, 2014). After the rationalization of both human labour and cognitive labour, we now enter a new phase where much of our society gets turned into data, and processed by autonomous, artificial entities.

The **digitalization** which turns our world into data is depicted in the figure (p. 272): each square represents an *object*, each triangle a *document* and each circle a *person*. Traditionally, all relations and interactions between any of these groups were *physical*. In our modern age, all such interactions are becoming *digitalized* step-by-step and produce data entering the data area. If we consider shopping, long ago, one could go to a store, fit some jeans, pay them and only the sales person (and the customer) would have a faint memory of who just bought which jeans. Nowadays, traces of security cameras, online search behavior on the store's website, Wi-Fi-tracking in the store, and the final payment, all generate a *data trace* of all interactions with the store and its products. A major consequence of that digitalization process is that a *permanent memory* of all those specific interactions is

---

[11] https://www.theguardian.com/news/2017/may/21/revealed-facebook-internal-rulebook-sex-terrorism-violence

[12] http://www.telegraph.co.uk/news/2017/06/16/facebook-using-artificial-intelligence-combat-terrorist-propaganda/

[13] https://www.theguardian.com/technology/2017/apr/07/google-to-display-fact-checking-labels-to-show-if-news-is-true-or-false

[14] https://www.forbes.com/sites/kalevleetaru/2017/02/17/how-twitters-new-censorship-tools-are-the-pandoras-box-moving-us-towards-the-end-of-free-speech/

[15] https://www.wired.com/2016/09/google-facebook-microsoft-tackle-ethics-ai/

[16] Wiener was, however, skeptical: "Again and again I have heard the statement that learning machines cannot subject us to any new dangers, because we can turn them off when we feel like it. But can we? To turn a machine off effectively, we must be in possession of information as to whether the danger point has come. The mere fact that we have made the machine does not guarantee that we shall have the proper information to do this." (N. Wiener (1948, 1961): Cybernetics, or control and communication in the animal and the machine).

[17] http://www.dailymail.co.uk/sciencetech/article-3624671/Google-s-AI-team-developing-big-red-button-switch-systems-pose-threat.html

[18] http://observer.com/2015/08/stephen-hawking-elon-musk-and-bill-gates-warn-about-artificial-intelligence/

[19] https://futureoflife.org/ai-open-letter/

[20] https://academic.oup.com/nsr/article/doi/10.1093/nsr/nwx045/3789514/Machine-learning-challenges-and-impact-an

[21] https://www.theguardian.com/technology/2016/sep/28/google-facebook-amazon-ibm-microsoft-partnership-on-ai-tech-firms

[22] https://www.nytimes.com/2016/11/02/technology/new-research-center-to-explore-ethics-of-artificial-intelligence.html?mcubz=1

[23] (In German) http://www.sueddeutsche.de/digital/alphabet-google-wird-allmaechtig-die-politik-schaut-hilflos-zu-1.3579711

[24] The start of this direction was only roughly ten years ago
The Petabyte Age https://www.wired.com/2008/06/pb-intro/ (Mitchell 2009)
Mining Our Reality http://www.cs.cmu.edu/~tom/pubs/Science2009_perspective.pdf (Anderson 2008)

[25] This phenomenon is called "the end of theory" since it breaks with standard scientific methodology.

[26] See for example Barcelona (http://www.smartcityexpo.com/barcelona) and other cities.

[27] See for example East-Germany's Stasi and the great movie about it 30 http://www.imdb.com/title/tt0405094/

[28] See the Rathenau Report on "Working in the Robot Society (2015) https://www.rathenau.nl/nl/node/766
The Rathenau Institute publishes many reports on the digital society and its implications, see https://www.rathenau.nl/nl/publicaties

"the cloud"
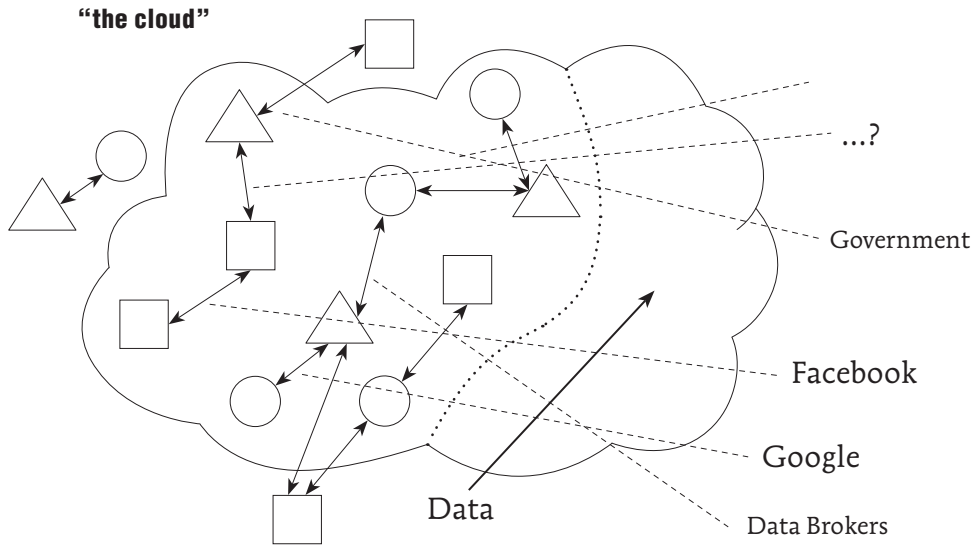
stored in a *cloud* and can never be forgotten. In addition, often this data is generated and governed by *private entities*. For example, Facebook governs a lot of our social interactions on their platform and keeps data about us, Google gathers everything that people do with its search engine, and Twitter keeps score of all our interactions via Tweets and others trace our love life (OkCupid, Tinder), our communication (Gmail, Twitter, WhatsApp) and our entertainment (Netflix). This data is to some extent owned by these companies, and whereas a long time ago interactions were physical, and no trace was kept, these modern platforms are *aimed* at gathering as much data as possible of all our interactions, and aimed at *retrieval* of that data (of all users combined) for purposes such as *profit* and *surveillance*.

Despite the focus on *data*, it is only a consumable for the entities that really change our world: algorithms. Algorithms are *computer programs* that autonomously utilize data in order to do something. This can be *sorting* names in a database, computing navigation instructions, or also organizing Facebook's news feed. The term algorithm[29] stands for any *finite procedure/recipe*, with *well-defined instructions* and which is *effective* in solving a problem. **Algorithmization** is the phenomenon where increasing numbers of tasks in society are carried out by intelligent algorithms. The field studying and creating such algorithms is AI[30] (McCarthy, 2007; Nilsson, 2010) which is seeing a recent explosion of advances, including breakthrough technologies such as reinforcement learning (Wiering and van Otterlo, 2012) and *deep learning*. AI's core is coming up with intelligent systems that *in some way* exhibit observable behavior for which some form of intelligence is required. Lately the focus is on *adaptive* AI, or *machine learning* (Domingos, 2015), which ranges from baking cookies[31] to driving autonomous cars by learning from popular computer games.[32] AI is rapidly becoming *the* driver for innovation (Stone et al., 2016).

The transformation into a digital society can thus be characterized by the two interrelating developments: *digitalization*, which turns once-physical interactions into digital *data*, and *algorithmization*, which amounts to increasing analysis and

utilization of that data by algorithms. The transformation's impact on archives (and libraries) is potentially huge. When it comes to digitalization, archives (and libraries) are in transformation. Collections are constantly being digitalized to provide wider public access to information, for example through the American project *Digital Public Library of America* (DPLA[33]) and the European counterpart *Europeana*.[34] They unlock massive amounts of archival data such as books, photographs and various documents. Initiatives such as the Google Books project are similar in terms of technology, but have different goals. Google Books[35] has a long history of battles[36] [37] [38] between a tech giant wanting to unlock *all* books written by mankind, for everyone, and author organizations that think that Google does not have the *right* to do that in this way. The ethical issues of access here are severe, since Google may want to push the idea being a *universal library* but many think this role should not be pursued by a commercial entity.

In general, libraries and archives (to some extent) have always struggled with their exact role, especially in the transformation to our digital age, with the novel aspects of born-digital records and books, and with the loss of being an information providing monopolist (Licklider, 1965; Herring, 2014, see also Anderson, 2011, p. 212) in the age of Google. Both Kallberg (2012) and Clement (2013) have investigated how the archival profession changes in our digital age, and how archivists think about that transformation. Paulus (2011) shows that the *lifecycle of information* of archives and libraries changes, and that, for example, a transformation is happening in which libraries may return to an ancient and medieval model of the library or archive as a site of both production and preservation. Cox (2007): "At last, archives have a real opportunity to abandon the role of gatekeeper and invite user participation, interaction, and knowledge-sharing." He continues: "What would happen if we could engage our users in defining and describing archival content and in communicating it to others? Is it possible that the analog archives tradition can learn from the movement of social media and social design? Some of the opportunities include diminishing the role of the archivist as gatekeeper, promoting participation and collaboration among users, and enriching the archives itself by tapping into the specialized and diverse knowledge of researchers".

The future of archives and libraries has many parallels with the development of information technology such as the internet. Noh (2015) describes several stages leading up to "library 4.0", which is where "technology will become one with users' lives" and which also features *3D printing, big data, cloud computing and augmented reality.* One can also digitalize interactions that were purely physical until very recently, for example using photocopiers[39] and (personal) cameras (Cox, 2007).

---

[29] https://en.wikipedia.org/wiki/Algorithm

[30] *Science*, special issue on how A.I. is transforming science http://science.sciencemag.org/content/357/6346/

[31] http://www.wired.co.uk/article/google-vizier-black-box-optimisation-machine-learning-cookies

[32] https://www.technologyreview.com/s/602317/self-driving-cars-can-learn-a-lot-by-playing-grand-theft-auto/

[33] http://dp.la

[34] http://www.europeana.eu

[35] http://books.google.com

[36] https://www.wired.com/2017/04/how-google-book-search-got-lost/

[37] https://www.theatlantic.com/technology/archive/2017/04/the-tragedy-of-google-books/523320/

[38] See https://www.theatlantic.com/technology/archive/2015/10/fair-use-transformative-leval-google-books/411058/ and https://www.wired.com/2017/04/how-google-book-search-got-lost/

[39] This also connects back to Eco's "restrictions" described earlier on being able to photocopy in a hostile library, but also to the ethical challenges concerning fairness when photocopying costs money.

Both Fernandez (2016) and van Otterlo (2016b) describe how AI can be employed to do, for example, recommendations based on access to items and user data. AI can also be employed for personal assistants (*agents*) implementing virtual reference desks (Liu, 2011), and to *optimize* library and archival processes. Many core archival processes can be *automated* but currently digitalization and algorithmization have only just begun.

Access to lots of information has been the dream of many visionaries, especially in the last century. Joseph Licklider (1965) predicted more than fifty years ago that humans by the year 2000 would invest in a kind of *intermedium*[40] which would provide access to the so-called *procognitive net*, containing all knowledge. Paul Otlet envisioned various automated ways to do knowledge classification and retrieval, and laid the foundation for the modern internet with his Mundaneum and universal decimal classification. In 1945 Vannevar Bush introduced the "Memex", resembling Otlet's "Mondotheque" (introduced around the same time), a machine in the form of a regular desk that used microfilm as the storage medium for collections of text, and which could provide access to knowledge. Otlet's version was more related to H.G Wells "World Brain" in the sense that it focused on "networked" knowledge, and targeted film, photographs and radio in addition to text. Wells, building on ideas on information retrieval in his early "A Modern Utopia" from 1905, introduced his "World Brain" in 1939 in a series of lectures, as an idea to make the whole human memory accessible to every individual. More recently Wilfred Lancaster wrote (1982, quoting Schiller 1977): "Ultimately, the signs point to a technology offering search capability at home or office terminals without the aid of librarian intermediaries who perform the searches." (p. 33-34). All these, and many more pioneers (see Borner (2010) and Wright (2014) for extensive overviews), envisioned forms of technology that would connect each individual to "all" knowledge, in the form of some "world encyclopaedia" and would make this knowledge retrievable by technology. In essence, our current world, with Google, Wikipedia, Twitter and smartphones, exhibits all that they were looking for.

The enthusiasm of these pioneers in "universally accessible" knowledge is echoed in today's Silicon Valley's technology push. Every day comes with new services, new technologies, new apps and new AI. That each person on earth, in principle, has access to the world's knowledge through a smartphone was just a start. Soon, algorithms will become the prime actor doing selection, ordering and description for many information-rich tasks. What Silicon Valley and the pioneers also have in common, at least until very recently, is their focus on the *possibilities* of novel technologies, and not on possible *(unintended) consequences*. Archivists, librarians and other information professionals have powerful roles as gatekeepers, and with great power comes great responsibility. If we are increasingly handing such tasks as *access to information* over to algorithms, or algivists, we need to look at the ethics of doing so. And, since human information professionals have been doing that for such a long time, it is interesting to see how they have handled moral issues in the next section.

(3) The Intended Archivist: Ethical Aspects of Archives

Taking practical action based on moral values is the domain of *ethics* (Laudon, 1995; Baase, 2013; Kizza, 2013). According to Kizza (2013) *morality* is "a set of rules for right conduct, a system used to modify and regulate our behavior." (p. 3).

It naturally has close ties to *law* since when a society deems certain moral values to be important, it can formalize such values in a law and set behavior that will uphold those values as a *norm*. Ethics typically is concerned with analysis of such norm-setting processes. Classic ethical questions are: "should we clone humans?", "is it sometimes allowed to kill people?" and "should we provide a base income in case robots take over most jobs?". As Laudon defines it (1995): "Ethics is about the decision making and actions of free human beings. When faced with alternative courses of action or alternative goals to pursue, ethics helps us to make the correct decision... Ethics is, above all, about what is good and what is evil, and how we come to make such judgments" (p. 34). I would summarize it as: if there are options what to do, then ethics is concerned with practical reasoning about "good" and "bad" actions. Important subsequent questions are then, *for whom* is something good or bad, and *by who's standards?* Different answers to those questions induce a variety of ethical reasoning frameworks, with two main dimensions. One is about *rules vs. consequences*: to find the right decision one may follow a religious rule like "thou shalt not steal", or look at the consequences and decide, for example ignoring a red light at night when there is no traffic. The second dimension deals with "for whom" something is good: the *individual*, or the *collective*. A well-known *collective consequentialist* framework is John Stuart Mills' *utilitarian* ethics, which is aimed at finding the decision that gives the best *result on average, for all*, and can be *unfair* to single individuals.

Traditional archives are filled with ethical issues. The archivist performs many core[41] archival operations that all involve ethical decisions. Archives are (just like libraries and museums, see Kirchhoff et al., 2008) "memory institutions".[42] Morris (2009): "Archives are records, regardless of format, created or received by a person or organization during the conduct of affairs and preserved because they contain information of continuing value." (p. 4). Archivists deal with the *selection* (acquisition, appraisal, accessioning, retention), *maintenance* (provenance, order, physical arrangements) and *description* (cataloguing, referencing) of sources. Access to the material in traditional archives involves physical access to the physical material. Because archivists are, in contrast[43] to e.g. librarians, highly involved in creating the order and descriptions of the archive, users are more dependent on the archivist when they want to access materials. Zastrow (2013): "The idiosyncratic and contextualized world of archives necessitates communication with the archivist." (p. 18). Physical access to archives and libraries has always appealed to our imagination, in *fiction, poetry* and *film* (Crawford, 2015). Exciting stories like *Indiana Jones* revolve around the idea of finding a lost archive and retrieving a valuable item. The nicest example of such a *physical hunt* for a book appears in Umberto Eco's (1980) *The Name of the Rose*, which features an evil librarian, a difficult book maze, and poisonous pages as *physical* barriers to access.

---

[40] In his words: "a capital investment in their intellectual Cadillac".

[41] https://www2.archivists.org/node/14804

[42] Kirchhoff et al. (2008, p252) cites Lorcan Dempsey (2000) as follows: "Archives, libraries and museums are memory institutions: they organize the European cultural and intellectual record. Their collections contain the memory of peoples, communities, institutions and individuals, the scientific and cultural heritage, and the products throughout time of our imagination, craft and learning. They join us to our ancestors and are our legacy to future generations. They are used by the child, the scholar, and the citizen, by the business person, the tourist and the learner. These in turn are creating the heritage of our future. Memory institutions contribute directly and indirectly to prosperity through support for learning, commerce, tourism, and personal fulfillment."

[43] https://www.quora.com/How-would-you-explain-the-difference-between-a-librarian-and-an-archivist

Archives have many stakeholders: users, donor organisation, archivist, and people occurring in documents. Any relation between them can cause dilemmas, and the archivist plays a pivotal role. A typical object of ethical study in this domain is *privacy* (Garoogian, 1991; Svensson et al., 2016). Preisig et al. (2014): "Librarians, archivists and other information workers had to face ethical conflicts and ethical dilemmas long before digital media and the Internet started to reshape the whole information sphere. Francis Bacon's aphorism knowledge is power (scientia potentia est) refers to the fact that limited access to information and restricted education were prerequisites of ruling elites in pre- and non-democratic societies." (p. 11). Many ethical dilemmas are about *access* but plenty others arise between archive stakeholders. For example, Preisig (2014) mentions that unlimited freedom of expression collides with protection from defamation: archives may contain information that, when published freely, could cause harm to individuals (rendering a conflict with the owner or the subject of the archival matter). Ferguson et al. (2016) introduce a list of 86 real-world ethical cases and cluster them by *dilemma*. Similar to Preisig et al. (2014) dilemma is the "privacy versus potential harm to individuals" but also included are "privacy versus organisational ethos or requirements" – where obligations to core customers were in conflict with the organisational interests, for example when a professor requests reading records of a student suspected of plagiarism – and "ethics versus law" – where librarians or archivists have a conflict between their ethical convictions and what they see as "unjust laws". An example of the latter was where the government instructed librarians not to buy books from a specific country. Next to data privacy, increased digitalization of archives and their use also creates challenges for *intellectual privacy* (Richards, 2015; van Otterlo, 2016a), which is the right of an individual to access and read whatever he *wants without interference or monitoring* and which is a fundamental requirement for intellectual growth, freedom of thought, and especially *autonomy*.

Access is the most important issue with ethical repercussions in archival practice. Danielson (1989): "Providing fair access to archives may appear to be a fundamentally simple operation, until one examines specific cases." (p. 53). It often comes down to balancing many interests of stakeholders, ranging from overzealous researchers who want to gain access to legitimately privileged papers, to archivists who disagree with institutional policies, and to donors who have difficulty relinquishing control over their papers. Danielson distinguishes three distinct cases concerning access: restricted collections, open collections, and the topic of fair access. The first two deal with ethical issues of various forms of (legal) access restrictions by donors because of privacy, or sensitive materials (e.g. government documents and possible war crimes). According to Danielson (1989): "Just as individuals are responding to a candid society with a renewed sense of privacy, so too are institutions showing a heightened awareness of security." (p. 59). Danielson's third case concerns *equal intellectual access*. In large archives it costs lots of work[44]

---

[44] An interesting case here is the one on Cybersyn, the socialist big-data-avant-la-lettre project from the seventies in Chile, which was extensively described by Eden Medina in her fascinating book "Cybernetic Revolutionaries" from 2011. In 2014 Evgeny Morozov wrote a piece in the New Yorker on the exact same project (http://www.newyorker.com/magazine/2014/10/13/planning-machine). This created some controversy because some people accused Morozov of plagiarism, and quite interestingly, his rebuttal consisted of showing photographs of his own extensive search efforts in the archives of Stafford Beer (the main person in the Cybersyn project). The issue was never fully resolved (http://leevinsel.com/blog/2014/10/11/an-unresolved-issue-evgeny-morozov-the-new-yorker-and-the-perils-of-highbrow-journalism).

and money to find interesting things. One idea to help users is to inform them when researchers are after similar items. Practically it is questionable whether this works. Danielson (1989) describes several hypothetical examples related to ethics. For example, do professors get priority over access to sources just because they are better researchers? Do fees for copy services influence the access, and should profit and non-profit making patrons pay the same fees? Should the judgment about the quality of a researcher make a difference when prioritizing access to particular still unpublished sources? And should ethical decisions be made when a journalist (who has a much faster publication medium) asks for the same information the archivist knows a researcher is working on?

The related study by Ferguson et al. (2016) lists five dilemmas where *access* to information comes into conflict with another important *value*. The first is *censorship*. For example, archives can contain materials about groups of people which some people might see as offensive, so a balance is needed between publishing information and protecting groups. The second is *privacy*: access to information and records of that access could be in conflict if the latter need to be shared, for example with authorities. The third dilemma concerns access and *intellectual property*. The example that is mentioned here is translating something into braille without copyright compliance. The fourth conflicting value consists of *social obligations*. This one is personal for the archivist: should he or she work (partially) for free in the context of budget costs, just to maintain the level of service? The last one concerns *organisational ethos or requirements*. Here the specific case was about making university theses publicly available (with pressure for "open access") even though this might jeopardise publication of the results.

Given the many ethical dilemmas in accessing archives, the big question is how do archivist know how to make the right choices? Several scholars all point to the use of so-called "**code-of-ethics**". A *code of ethics* formalizes rules, guidelines, canons, advisories and more for the members of a particular profession. Well-known examples are the *ten commandments*[45] of the Christian bible and Asimov's *three laws of robotics*[46]. Another influential code is the *universal declaration of human rights* which deals[47] with fundamental ethics of human life. Usually codes of ethics are used by high-visibility institutions and big corporations[48], but in principle any profession could define one. The main *objectives* of a code of ethics are five-fold:

- **Disciplinary**: to enforce professionalism and the integrity of its members, possibly with *penalties*.
- **Advisory**: to offer members advice when difficult ethical decisions need to be made, professionally.
- **Educational**: to educate new members and show them the do's and don'ts of the profession.
- **Inspirational**: to (indirectly) inspire members to "do the right thing".
- **Publicity**: to show *externally* that a profession and its members have clear values and moral behavior.

---

[45] https://en.wikipedia.org/wiki/Ten_Commandments

[46] https://en.wikipedia.org/wiki/Three_Laws_of_Robotics

[47] In previous work (van Otterlo 2014b) I analyzed this code and found several necessary alterations needed for the digital age. Recently more interest in such issues has risen, due to advances in AI and robotics (Van Est, R. and Gerritsen, J. 2017).

[48] See for example one by IKEA (http://www.ikea.com/ms/en_JP/about_ikea/our_responsibility/iway/index. html), by Sony (https://www.sony.net/SonyInfo/csr_report/compliance/index3.html) and McDonalds (http://corporate.mcdonalds.com/mcd/investors/corporate-governance/codes-of-conduct.html)

Codes of ethics can be *prescriptive* (prescribe the do's and don'ts) or *aspirational* (only specify *ideal results*). Ferguson et al. (2016) note that they are an important tool for archivists, yet not always sufficient, especially not when there are *conflicts* between rules and values.

Archival codes[49] of ethics have a history. The first dates from 1955, from the Society of American Archivists (SAA). It (SAA 1955) is fairly compact and states things like:

> "The Archivist should endeavour to promote access to records to the fullest extent consistent with the public interest, but he should carefully observe any proper restrictions on the use of records".

Similar statements come from the Universal Declaration on Archives (ICA-DL 2011):

> "Archives are made accessible to everyone, while respecting the pertinent laws and the rights of individuals, creators, owners and users".
>
> "The Archivist should respond courteously and with a spirit of helpfulness to reference requests."
>
> "The Archivist should not profit from any commercial exploitation of the records in his custody."

Later (SAA 1992) it includes:

> "It is not sufficient for archivists to hold and preserve materials: they also facilitate the use of their collections and make them known."

This amounts to the preservation, use and publicity aspects of the archive. It also contains:

> "Archivists endeavour to inform users of parallel research by others using the same materials, and, if the individuals concerned agree, supply each name to the other party."

This refers to a dilemma I have discussed.

The final commentary of the code states something about potential conflicts:

> "When there are apparent conflicts between such goals and either the policies of some institutions or the practices of some archivists, all interested parties should refer to this code of ethics and the judgment of experienced archivists."

The most recent version (SAA 2012) features additional *core values*, which represent what the archivists *believe* while the code itself represents a framework for the archivists' *behavior*. This division is intuitive and could be a way to solve some ethical dilemmas, for example by a utilitarian analysis weighing in more factors. For access it expresses the value that it is essential in personal, academic, business and government settings, and use of records should be welcomed. Later in the code of ethics itself this value is translated into "minimize restrictions and maximize ease of access".

Ethical codes, especially when they have consequences when misbehaving, cause fewer discipline problems among members (Kizza, 2016, p. 50). However, some codes of conduct can be *non-committal*. Morris[50] calls for an *enforceable* code of ethics, just like legal and medical professions are governed by codes of ethics which carry the force of law. Violations would then be subject to sanctions including loss of license and civil and criminal liabilities. Formalizing ethical codes though, has one main purpose: to formalize *how humans should behave*, in this case in the archival profession. I call this the intended archivist; how he is supposed to think, feel and act professionally based on *human values* and *human behavior*. By formalizing it in a code it becomes *transparent* and can be communicated to peers, users, donor organisations and the general public.

(4) The Ethics of Algorithms

For algorithms, ethical analysis has only started recently resulting in the multidisciplinary field of *ethics of algorithms* (see for pointers: Lichocki et al., 2011; van Otterlo, 2013,2014a,2014b,2016a; Medina, 2015; Mittelstadt et al., 2016). People often associate with algorithms properties such as *infallible, exact*, and especially: *objective*. Because computer-based algorithms are based on logic and statistics people tend to think that because of that algorithms are objective and fair, since they can compute the *best answers* given the data. While some of this may be true, in general algorithms are far from objective: they are heavily *biased* (Bozdag, 2013; van Otterlo, 2013). Consider for example[51] (part of) a simple algorithm for a bank, specifying that "IF sex = female AND age > 60 THEN decision = no-life-insurance-policy". Now this algorithm is perfectly mathematical, and exact, and it thoroughly computes from personal data whether somebody is eligible for a life insurance policy. However, from a human point of view, it is far from "objective", or "fair" since it discriminates against women above 60 years old. Its decisions are *biased* and it discriminates, in plain sight. To make things worse, we can also imagine a second algorithm which is specified as "IF f(sex) * g(age) > 3.78 THEN decision = no-life-insurance-policy", and let us assume it makes exactly the same decisions as the first. A problem here is that this algorithm discriminates too, but it is hard to see from its description because we do not know what the functions f() and g() do, and also not why there is a threshold of exactly 3.78. Maybe these aspects have been *learned from data* which would require us to have a look at the data and learning process to form an opinion about the algorithm's objectiveness. In general, algorithms *are* biased in many ways (Bozdag, 2013), for example by the data, by learning procedures, by programmers who make choices, by technological constraints and many other reasons. This immediately requires us to form an opinion about algorithms and whether they do *the right thing*, which again brings us back to ethical reasoning.

Characterizing the ethics of algorithms is hard since algorithms and potential consequences are so diverse, and situations may change over time. Mittelstadt et al. (2016) define *concerns* about how algorithms transform data into decisions, which are then coupled with typical ethical issues. The core operations of an algorithm are: 1) it turns *data* into *evidence* which can be a probabilistic prediction, a yes-no decision, or some other *conclusion*, and 2) it uses the evidence to *trigger and motivate an action* based on the data. For example, an algorithm for bank loans could take personal data of someone and produce a credit-score of 12, which then could trigger an action to approve a particular mortgage. For the first step three general concerns

---

[49] Many other related codes exist, for example by the Dutch royal association for archivists (KVAN)(1) and the professional charter for librarians in public libraries (PL 1993)(2), and codes by the American library organization (ALA)(3), the International Federation of Library Associations and Institutions (IFLA)(4) and the International Council of Museums (ICOM)(5) code of ethics for museums. Although libraries do have different activities, core values are shared with archivists, which can be seen in the similarities with library values concerning access. Occasionally separate codes are made with respect to specific aspects such as privacy, for example as was done recently by IFLA in 2015 (6). See: (1) http://kvan.nl/images/PDF/Beroepscode_voor_Archivarissen.pdf; (2) http://www.ifla.org/files/assets/faife/codesofethics/netherlands.pdf; (3) http://www.ala.org/advocacy/proethics/codeofethics/codeethics; (4) http://www.ifla.org/news/ifla-code-of-ethics-for-librarians-and-other-information-workers-full-version; (5) http://icom.museum/fileadmin/user_upload/pdf/Codes/code_ethics2013_eng.pdf; (6) https://www.ifla.org/node/9803

[50] http://slanynews.blogspot.nl/2010/09/enforceable-code-of-ethics-why.html

[51] Birkbak and Carlsen (2016) elegantly show in a toy experiment how bias that is explicitly put in (the code of) a ranking algorithm causes different results, exemplifying how implementation choices change algorithm outcomes. As bias, they use intuitive operationalizations of the company mottos of Google, Facebook and Twitter.

can be defined. First, the evidence may be *inconclusive*. For example, when an algorithm predicts that I am a terrorist with 43.4 percent probability, what does it mean? Second, evidence may be *inscrutable* and not open for inspection which is often the case for a *no-fly list* decision. Third, evidence can be *misguided*, meaning that the underlying data is incomplete or unreliable. Actions, decided upon evidence, may have problems too, since they can be unfair, e.g. *discriminatory*. In addition, they can have *transformative* effects, for example that they change people's behavior which can happen when Facebook orders your personal news feed. These concerns then lead to typical patterns with ethical implications. For example, transformative effects can lead to *loss of autonomy* when a search engine manipulates you with advertisements, inconclusive evidence can lead to *unjustified actions*, and inscrutable evidence can lead to *opacity*. Overall, many concerns lead to a *loss of privacy*, and for any *algorithmic decision-making* situation attributing responsibility for the decisions can be quite complicated.

As a complement to this taxonomy, I developed[52] another way to look at the potential (ethical) impact of algorithms, now ordered by *what the algorithm can do*, or in general terms their *level of autonomy*. This results in five broad algorithm classes which have clearly defined *capabilities* and corresponding ethical issues.

Algorithms that interpret

The first type consists of *algorithms that reason, infer and search.* These algorithms can be quite complex in what they do, but they all compute answers based on data *as it is*. The more complex they are, the more information they can extract from that data. Examples include translation[53] and spatial language understanding[54] but also poetry generation.[55] Visual information processing now includes examples in recognizing[56] *what is on a picture*, evaluating picture's *aesthetics*[57], generating 3D face[58] models, *augmented reality* with IKEA[59] furniture and even recognizing kids in Halloween[60] costumes by Google's autonomous cars. The interpretation of sound includes better-than-human speech recognition[61], lip reading[62], and real-time Skype translations.[63] General *data science* can for example be used to *infer*[64] when people get into love relations. Ethical concerns about such algorithms are typically about *privacy* since more ways become available to interpret and link more kinds of data. A second member of this group are *search* algorithms like Google. They not only rank and filter information, but they increasingly so use *knowledge* and *learning* to understand what the user wants (Metz, 2016a). Search engines also try to *answer* queries like "how high is the Eiffel tower" instead of delivering source documents.

---

[52] In the context of my course on the ethics of algorithms, see http://martijnvanotterlo.nl/teaching.html
[53] https://translate.google.com/?hl=nl
[54] https://www.wordseye.com/
[55] http://www.wired.co.uk/article/google-artificial-intelligence-poetry
[56] https://www.theverge.com/2017/6/15/15807096/google-mobile-ai-mobilenets-neural-networks
[57] https://petapixel.com/2016/10/08/keegan-online-photo-coach-critiques-photos/
[58] https://petapixel.com/2017/09/20/ai-tool-creates-3d-portrait-single-photo/
[59] IKEA augmented reality https://www.youtube.com/watch?v=UudV1VdFtuQ
[60] http://www.dailymail.co.uk/sciencetech/article-3301013/Google-teaches-self-driving-cars-drive-slowly-children-dressed-up.html
[61] https://www.technologyreview.com/s/544651/baidus-deep-learning-system-rivals-people-at-speech-recognition/
[62] https://www.technologyreview.com/s/602949/ai-has-beaten-humans-at-lip-reading/
[63] https://futurism.com/skype-can-now-translate-your-voice-calls-into-10-different-languages-in-real-time/
[64] https://www.facebook.com/notes/facebook-data-science/the-formation-of-love/10152064609253859/

The ethical issues with search engines are typically about the transformative effects they have on user *autonomy*, because of their enormous power (Granka, 2010; van Otterlo, 2016a). Search engines are key *gatekeepers* and influence the minds of billions of people every day. They have been shown to be capable of influencing[65] elections (Anthes, 2016), which is a serious ethical problem. Answering queries is an important issue too in so-called *conversational agents* and *social bots* (Ferrara et al., 2016). Social bots can influence discussion on forums, or act as genuine users on platforms such as Twitter. An ethical issue is that bots could be used for malicious[66] purposes, such as steering a debate towards a particular outcome, or providing false support for election candidates. This raises threats for autonomy again as a transformative effect. A second type of conversational agent are the *voice-controlled assistants*[67] such as *Cortana, Siri* and *Alexa*, which perform tasks like agenda keeping, creating shopping list, and *answering questions*. Assistants are increasingly used, especially in China[68], and have already appeared[69] in legal[70] situations (as a "witness").

Algorithms that learn

The second class of algorithms goes beyond the first and can *learn*, and *find generalized patterns* in the data. These *inductive* algorithms perform statistical inference to derive patterns, models, rules, profiles, clusters and other *aggregated knowledge fragments* that allow for statistical predictions of properties that may not be explicitly in the data. Overall, these are typically *adaptive* versions of the inference algorithms I have discussed, i.e. search engines typically adapt over time, and algorithms that interpret text, images and sound are often trained on such data. Applications range from predicting sounds for video[71], to training self-driving cars using video game data[72], even to predicting social security numbers.[73] Once algorithms start to learn (Domingos, 2015; Jordan and Mitchell, 2015) from data concerns about inconclusive evidence are justified because most methods use *statistical predictions*. In addition, outcomes may change over time with the data, making outcomes unstable. Most powerful contemporary learning algorithms, such as *deep learning*[74][75], are purely statistical algorithms and very much like black boxes, which entails they are non-transparent and the evidence they produce inscrutable (with some exceptions[76]). When algorithms are used for *profiling* and *personalization* (van Otterlo, 2013; De Hert and Lammerant, 2016), something that happens everywhere on the internet, algorithms influence the user's choices and therefore affect his *autonomy of choice*. If profiles are learned from data, algorithms

---

[65] https://algorithmwatch.org/en/watching-the-watchers-epstein-and-robertsons-search-engine-manipulation-effect/

[66] A funny example of a malfunctioning bot: https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

[67] http://www.businessinsider.com/siri-vs-google-assistant-cortana-alexa-2016-11?international=true&r=US&IR=T

[68] https://www.technologyreview.com/s/608841/why-500-million-people-in-china-are-talking-to-this-ai/

[69] See also the hilarious Southpark episode on these assistants: http://www.ibtimes.com/south-park-season-premiere-sets-amazon-echo-google-home-speakers-2590169

[70] https://www.wired.com/2017/02/murder-case-tests-alexas-devotion-privacy/

[71] https://www.engadget.com/2016/06/13/machines-can-generate-sound-effects-that-fool-humans/

[72] https://www.youtube.com/watch?v=JGAIfWG2MQQ

[73] https://www.wired.com/2009/07/predictingssn/

[74] https://www.wired.com/2017/04/googles-dueling-neural-networks-spar-get-smarter-no-humans-required/

[75] https://machinelearningmastery.com/inspirational-applications-deep-learning/

[76] http://www.sciencemag.org/news/2017/07/how-ai-detectives-are-cracking-open-black-box-deep-learning

typically learn statistical models *from many* users and *apply them to a single user*. This may render inconclusive evidence which may be right on *average* but not for that single individual. A new *privacy* risk of learning algorithms is that they can also reveal *new* knowledge (van Otterlo, 2013; Schwartz et al., 2013; Youyou et al., 2015; Kosinski et al., 2013), predicting personal traits from language use, Facebook like's or just a photo.[77] Such algorithms obviously have effects on privacy, but certainly also transformative effects related to *autonomy*.

A more general consequence of adaptive algorithms is that we move in the direction of "the end of code" (Tanz, 2016). In the near future, increasingly many algorithmic decision-making tasks will be learned from data, instead of hardcoded by programmers. This has consequences for society, and for people, who will more often be assigned the role of *trainer*, instead of *programmer*.

Algorithms that optimize

The third class of algorithms consists of algorithms that *optimize, incorporate feedback*, and *experiment*. These typically employ *reward functions* that represent what are *good outcomes*, which can be, for example, a sale in a web shop, or obtaining a new member on a social network. *Reward definitions* tell an algorithm what is *important* to focus on. For example, advertising algorithms on webpages get +1 reward for each time a user clicks on an offer. Optimization algorithms will, based on all that is known about statistical aspects and based on all data about a problem, compute the *best* expected solution. The most prominent system currently comes from Google's DeepMind. It combines reasoning, learning and optimization, beat the world best Go player (Metz, 2016b) and is now tackling the complex computer game Starcraft-2.[78] Optimization algorithms feature two kinds of rewards. One is used by the algorithm to optimize and represents clicks, sales, or other things which are *valuable*. The other type are rewards *for users* (e.g. a sale), with the goal of *nudging*[79] them into doing something (e.g. buying something). Manipulating users' behavior obviously has transformative effects on autonomy. Worse, just like learning algorithms, optimization works well *on average* and could deliver nudges to the wrong users too, which would make the outcomes discriminating and unfair. Optimization algorithms typically *iterate* the optimizations by **experimenting** with particular decisions, through *interactions with the problem* (see Wiering and van Otterlo, 2012). A good example are algorithms that determine the advertisements on the web: they can "try out" (experiment) with various advertisements for individual users, and use the feedback (clicking behavior) of individuals to *optimize* advertisement placings. So, instead of a one-pass optimization, it becomes an *experimentation loop* in which data is collected, decisions are made, feedback and new data is collected, and so on. Platforms with large user bases are ideal laboratories for experimentation. For example, Netflix experiments with user suggestions to optimize their rewards which are related to how much is being watched (Gomez-Uribe and Hunt, 2015). Optimization algorithms are generally used to rank things or people. In the *ranked society* in which we now live everything gets ranked, with examples such as Yelp, Amazon, Facebook (likes), TripAdvisor, Tinder (swiping) and OkCupid, all to find "the best" restaurant, lover, holiday trip, or book. Also in our work life, ranking and scoring becomes the norm (called: *workplace monitoring*[80]). The ultimate example is China's 2020 plan (Chin and Wong, 2016) to rank everyone in society to find out "how good a citizen are you". Scores are computed from many things ranging from school results to behavior on

social media, to credit score, and combined into one overall score. The higher that score, the more privileges the citizen gets (from easier car rental and bank loans, to visa to other countries). The ethics of experimentation has many aspects (Puschmann and Bozdag, 2014). Most important here are the choice of reward function (who decides has great power) and the fact that (especially on the internet) we often do not know we are part of an experiment, and maybe we need new forms of *consent*.

Physical manifestations

A fourth class of algorithms concerns *physical manifestations* such as *robots* and *sensors* (internet-of-things). These algorithms go beyond the digital world and have *physical presence* and *agency* in our physical world, which may jeopardize human safety. A first manifestation is the *internet-of-things* (Ng and Wakenshaw, 2017) in which many appliances and gadgets get connected and where increasingly *sensors* are being placed everywhere[81], creating data traces of once physical activities. The *programmable world* (Wasik, 2013) will feature all digital (and intelligent) items around us as being *one giant computer* (or: algorithm) that can assist us and manipulate us. For example, if your car and refrigerator and microwave could work together, they could – with the right predictions on the weather, your driving mood and speed, and possible traffic jams – have your diner perfectly cooked and warm the moment you get home from work. The ubiquity of such systems will raise ethical issues since they will be influential, but often unnoticeable. Also, privacy concerns are raised. A similar big development will be physical *robots*[82] in our society. "A robot is a constructed system that displays both physical and mental agency, but is not alive in the biological sense" (Richards and Smart, 2016). Many types of robots exist, ranging from simple vacuum cleaners, to *humanoids* (with human-like appearance[83] [84]) to robots capable of manipulating their physical environments in hospital or manufacturing situations. Robots are not yet part of our daily lives, but the literature on the ethics of robots is rich (Lichocki et al. 2011; Smart and Richards, 2016). Steinert (2014) frames the ethics of robots into four main[85] categories: robots as *tools* (or instruments), robots as *recipients of moral behavior,* robots as *moral actors*, and robots as *part of society*. The difference between the first and the latter two is mainly one of *responsibility*. The introduction of increasing numbers of robotic agent in society (the fourth type) will also have socio-economic consequences we can only partially imagine, most obviously for *work* which will[86] increasingly being taken (or not[87]) over by robots (Ford, 2013). Robots are also expected to have (ethical) impact on things like law enforcement, the military, traffic (Kirkpatrick, 2015), healthcare and even prostitution (Richardson, 2016).

---

[77] https://www.theguardian.com/technology/2017/sep/12/artificial-intelligence-face-recognition-michal-kosinski

[78] https://deepmind.com/blog/deepmind-and-blizzard-release-starcraft-ii-ai-research-environment/

[79] https://en.wikipedia.org/wiki/Behavioural_Insights_Team

[80] https://harpers.org/archive/2015/03/the-spy-who-fired-me/

[81] https://www.iamexpat.nl/lifestyle/lifestyle-news/hidden-cameras-dutch-advertisement-billboards-ns-train-stations-can-see-you

[82] https://en.wikipedia.org/wiki/R.U.R.

[83] https://en.wikipedia.org/wiki/Uncanny_valley

[84] https://www.wired.com/2017/04/robots-arent-human-make/

[85] The article also includes a fifth type which refers to the influence of robots on ethics itself (meta-ethics).

[86] https://www.wired.com/brandlab/2015/04/rise-machines-future-lots-robots-jobs-humans/

[87] https://www.wired.com/2017/08/robots-will-not-take-your-job/

Superintelligence

The fifth class of algorithms goes beyond the algorithms as we know them now (digital or in physical form) all the way to *superintelligent* algorithms, which surpass our human-level intelligence. Once we have reached that point, questions of *conscience* and *moral decisions*, and with that *responsibility* of algorithms, will play a role. Most of this discussion falls beyond the scope of this text. A general remark is that the more intelligent, autonomous or conscience an algorithm will become, the more moral values will be attributed to it, and the more ethical reasoning and behavior will be expected of it. However, as Richards and Smart (2016) elegantly show using the *android fallacy* it will take still a long time before robots are even capable of deserving that. According to many scholars, a so-called (*technological*) *singularity* (Vinge, 1993; Shanahan, 2015) will come, which is[88] "the hypothesis that the invention of artificial superintelligence will abruptly trigger runaway technological growth, resulting in unfathomable changes to human civilization". For some already the point of getting algorithms to become "smarter" than humans (whatever that may mean) will trigger an explosion of unstoppable AI growth that could dominate the human race entirely even. Ethical concerns about such algorithms are discussed by Bostrom and Yudkowsky (2011) and many other people, like Kurzweil.[89] Many straightforward ethical concerns are about whether machines will overpower us, whether they still need "us", and what it *means* to be human in a society dominated by machines (see Shanahan, 2015 for some pointers).

These five groups of algorithms show the many sides of the ethics of algorithms. Depending on the type of algorithm, task, setting and data, many kinds of ethical issues arise that must be addressed.

(5) Towards the Intentional Archivist

Algorithmic versions of virtually all current professions will appear, eventually. The basic, *human*, question is how to ensure that all these algorithms respect our human values. In this section I will sketch the considerations in ensuring algorithms like Paul, the algivist from the scenario at the beginning of this essay, will have the right moral behavior if we actually build them.

Solving ethical issues using AI

The previous section has described many potential ethical issues and they would all apply to algivists, but so far not many effective *solutions* exist. Literature on *governance* of algorithms (Diakopolous, 2016) focuses on *transparency* and *human involvement*, and on *making algorithmic presence known*. A challenge is that so far algorithms are largely *unregulated* (van der Sloot et al., 2016). However, there are laws and rules for *data*, such as the *data protection act* (DPA; Dutch: WBP[90]) from 1998. In 2018 new European regulation will take effect as a replacement of the directive of[91] 1995 in the form of the *general data protection regulation* (GDPR[92]) which will cover several forms of algorithmic decision making (see also Mittelstadt et al., 2016). Outside the law, solutions include *privacy-by-design*, and *encryption*.

---

[88] https://en.wikipedia.org/wiki/Technological_singularity
[89] https://en.wikipedia.org/wiki/The_Singularity_Is_Near
[90] http://wetten.overheid.nl/BWBR0011468
[91] https://www.autoriteitpersoonsgegevens.nl/nl/onderwerpen/europese-privacywetgeving/algemene-verordening-gegevensbescherming
[92] General Data Protection Regulation (GDPR) http://www.eugdpr.org/more-resources-1.html

Individual users can often protect their privacy to some extent by using privacy-friendlier software or services. A solution shared by many is *data minimization* (see e.g. Medina, 2015): only gather data that is really necessary. Another set of solutions is *obfuscation* (Brunton and Nissenbaum, 2013) in which users deliberately sabotage algorithmic systems.

An alternative though, is to *employ AI itself*. That is, one can utilize the same power of algorithms to deal with ethical issues. For example, recent advances in machine learning remove discriminatory biases by adapting training methods, or implement *privacy-aware* techniques. Etzioni and Etzioni (2016) propose general *AI Guardians* to help us cope with the government algorithms. Since AI systems more and more become *opaque* (black box), *adaptive* (using ML) and *autonomous*, it becomes undoable *for humans* to check what they are doing and AI systems can do that for us. AI guardians are *oversight* systems using AI technology, and come in various sorts: *interrogators* can investigate e.g. a drone crash, and a *monitor* can keep an eye on other AI systems, or even *enforce* compliance with the law. A special type is the *ethics bot* which is concerned with ensuring that the operational AI systems *obey ethical norms*. These norms can be set by the individual, but can also come from a community. An ethics bot could guide another operational AI system, for example to ensure a financial AI system only invests in socially responsible corporations.

Learning the right values

 Ethics bots will have to *learn* moral preferences, either by explicit instruction or from observed behavior. An intuitive idea would be to let algivists learn their moral behavior, for example, from watching a human archivist do their work. AI has developed many ways to do that, for example using *imitation*, or *learning from demonstrations*, however it is not that simple. A key challenge is *generalization*: which parts of the task need to be imitated exactly, and which not? "We're always learning from experience by seeing some examples and then applying them to situations that we've never seen before. A single frightening growl or bark may lead a baby to fear all dogs of similar size – or, even animals of every kind. How do we make generalizations from fragmentary bits of evidence? A dog of mine was once hit by a car, and it never went down the same street again – but it never stopped chasing cars on other streets." (Minsky, 1985, Society of Mind, Section 19.8). Based on the advances I described in the previous sections, AI would be capable of recognizing and interpreting the actions of a human archivist in action, and also replicating them in a robotic body, but it would still be a challenge to *do* learn how to sort out documents and to appraise the documents in the boxes, but to not learn how to scratch a nose, or fingertap while waiting for the printer to finish.

An effective alternative is to *learn the underlying reward function*. As we know from optimization algorithms, a reward function determines what is important and what not. Now assume the algivist could learn the reward function according to which the archivist does his job. In that case, the algivist would be able to replicate the archivist's behavior, including all the right ethical decisions. The technical term for this type of learning is *inverse reinforcement learning* (Wiering and van Otterlo, 2012) which is based on solid theories for behavior learning. For specialized tasks, especially in robotics, many successful applications exist. Equally so, it could form the basis for AI systems that *act* in alignment with human goals and values, which is an interesting option for ethical algivists. The core challenge then is how to learn these *human* values, sometimes framed as *the value learning problem* (Soares, 2015).

The challenge is that human values are typically difficult to learn, since they can be based on complex mental processes, can be working on multiple timescales, can be difficult to put on one value scale, can involve both intuition and reasoning and may involve other interactions such as *signalling* and *trust-building*. Furthermore, they require *ontological agreement* between human and machine: do they see the world in the same way? Many of these problems are shared with technical AI work (e.g. computer vision) but for use in ethical systems much more work is needed.

Against learning from scratch

The value learning problem is difficult for many reasons. In addition, any type of purely statistical learning procedure faces other difficulties related to opacity and the limited possibilities to employ *knowledge* one might have about a domain. However, there are machine learning techniques that allow for the *insertion* of knowledge as a *bias* for learning, and the *extraction* of learnt knowledge after learning. Consider the robot learning technique by Moldovan et al. (2012) where a robot needs to learn from demonstration how physical objects are to be manipulated and how they behave when manipulating. Without any prior knowledge, the robot would have quite a challenging learning problem, mapping the pixels of its cameras all the way to motor commands in its hands. Instead, by adding some common-sense knowledge about the world, like "if you move object A, and object B is far away, then you can safely assume B will not be affected", or "if you want to manipulate an object, you can either push, tap, or grab". This type of knowledge will make the learning problem easier and at the same time it *focuses* (or: *biases*) the learning efforts on the things that really matter. Other, general common-sense knowledge could also help in choosing the right behavior (based on a reward function) such as "green objects are typically heavy", and "one cannot place an object on a ball-shaped object". In machine learning we call this kind of bias *declarative*, since it is knowledge that can be *explicitly* used, stored, and "looked at". Declarative models have been used before in ethical reasoning in AI (Anderson and Anderson, 2007) and other ethical studies (van Otterlo, 2014a).

In order for inserting knowledge to work, we need to solve the ontological issue: knowledge should be at the right level and meanings should mean the "same" for AI and for humans. To bridge AI and human (cognitive) thinking, the *rational agent* view is a suitable view. In AI, a rational agent is "one that acts so as to achieve the best outcome or, when there is uncertainty, the best expected outcome" (Russell and Norvig, 2009). In cognitive science we can take the **intentional stance** view introduced by Daniel Dennett (2013). The intentional stance sees entities as *rational agents* having *mental notions* such as *beliefs, goals* and *desires*. Using this viewpoint, we assume the agent takes into account such beliefs and desires to *optimize* its behavior. For people this is the most intuitive form of description of other people's behavior. But, it is also common to use it to talk about algorithms: I can say that Google *believes* I like Lego and therefore it *desires* to feed me advertisements about it and sets a *goal* to prioritize websites referring to Lego. I can also say that Google *believes* that I *want* pizza when I enter "food" as a query since it *knows* from my profile it is my favourite food.

Code of ethics as a moral contract between humans and machines

Coming back to the *archivist singularity* mentioned in the introduction, I propose a simple strategy to construct **Paul, the Intentional Algivist** as a robotic,

algorithmic agent for the archives that has moral principles just like human archivists. What could be better *declarative, human* knowledge about ethical values in the archival domain than the previously discussed archival *codes of ethics?* Indeed, these hold general consensus ideas on how an archivist should behave ethically, dealing with issues such as privacy, access, and fair use of the archive. In addition, they are full of *intentional* descriptions, see for example: "The Archivist should endeavour to promote access to records to the fullest extent consistent with the public interest, but he should carefully observe any proper restrictions on the use of records". This is clearly a bias on how the algivist should behave and it contains intentional constructs such as a goal, a desire and several (implicit) beliefs. Codes of ethics are solid knowledge bases of the most important ethical *guidelines* for the profession, and typically they are defined to be *transparent, human-readable* and *public*. Using codes of ethics as a knowledge bias in adaptive algivists that learn ethical behavior is natural, since it merely *translates* (through the rational agent connection) an ethical code that was designed as a bias for human behavior, and uses that as a guide or constraint, or: *as a moral contract between man and machine*. I see a practical way to go in which an algivist is endowed with the ethical values contained in the code of ethics, after which it observes human archivists at work to *fine-tune* its behavior based on their example. Human archivists will slowly transform into *trainers and coaches of algivists*: the more advanced algivists become, the more humans will *guide* them and leave the archival work to them. But, before this happens, much still needs to be done, both by AI researchers as well as by archivists themselves.

What does the field of AI need to do?

AI needs to keep on progressing as always, but more research is needed on several aspects specifically. Language understanding and formalization of human (common-sense) knowledge needs to be improved to translate codes of ethics automatically in forms that the algivist can use for acting, and for reasoning. We know that even the impossible Roadrunner cartoon logic has at some point been formalized (McCartney and Anderson, 1996), so nothing is impossible. Furthermore, robotic skills need to improve a lot. Manipulation skills are somewhat sufficient for laboratory conditions (e.g. Moldovan et al., 2012) and there has been some progress in – for archivists, related – environments such as libraries[93], but obtaining general movement and object manipulation skills in *any* physical archive will take enormous efforts still. Once parts of the archive have been made digital, many of the archival selection, ordering and description tasks can be handled well, although also there much improvement is possible in the *semantic understanding* of documents, images, and other items.

What do archivists need to do?

Archivists will need to assist AI researchers as experts in archives, and they need to decide at least two things.

- The ethics of choosing THE code of ethics: The core idea is to inject ethical codes into machines. Out of the many possible versions, which one should be picked? And who decides upon that? Archivists, committees of experts, programmers, or more general democratic methods? For this to work, we may also need to investigate more which kinds of values hold in professions as held by archivists and librarians.

---

[93] https://phys.org/news/2016-06-automated-robot-scans-library-shelves.html

- Who approves algivists? Depending on the impact of algivists on the daily life of people, we may need regulations concerning their *use*, similar to regulations concerning autonomous cars. In analogy with medicine, we may need to think about formal *approval* procedures, as a kind of "FDA approval for algivists" (Tutt, 2017) where algivists first need to be tested in laboratories before they can work in archives.

(6) Conclusions

In this essay I have done several things. First, I have described a plausible, yet still fictive, future of algivists as algorithmic archivists. Second, I have given the reader an extensive view on the new field of ethics of algorithms which is concerned with the societal impact of intelligent algorithms. Third, I have introduced how ethical thinking has been formalized in archival practices using codes of ethics. Fourth, and most importantly, I have sketched how human codes of ethics could be employed to build algorithmic algivists that will obey our human ethical values, thereby moving from the *intended archivist* to the *intentional algivist*. Future research on AI, ethics and archives will bring us the advances algivists promise, but hopefully also gives us tools to maintain high moral standards when incorporating these agents into our lives. At least two sets of questions remain for archivists to answer and for others to study. The first is how the new role of *trainer* or *coach* will have an impact on the profession. Instead of appraising documents, they will now "appraise" the behavior of the algivist. How many trainers are needed at some point? How will they become a *certified* trainer? And will algivists stay a separate profession, or will they merge with other information service machines? The second set of questions is about the long-time future. When archival practices will be changed drastically with the introduction of algivists, ethical codes may need to be updated to reflect new social norms between humans and machines. Who decides when that time comes, and who decides what is to be changed? Maybe employing algivists also requires us to rethink ethical concepts over time (see Steinert, 2014 about meta-ethics). Time will tell.

Literature

ICA-DL (2011). *Universal Declaration on Archives*. Retrieved from http://www.ica.org/en/universal-declaration-archives

SAA (1955). *Society of American Archivists Code of Ethics* (1948-1965). Retrieved from https://historyhub.history.gov/docs/DOC-1031

SAA (1992). *Society of American Archivists Code of Ethics*. Retrieved from http://ethics.iit.edu/ecodes/node/4559

SAA (2012). *Society of American Archivists Code of Ethics*. Retrieved from http://www2.archivists.org/statements/saa-core-values-statement-and-code-of-ethics

Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired Magazine*. Retrieved from https://www.wired.com/2008/06/pb-theory/

Anderson, M., & Anderson, S.L. (2007). Machine Ethics: Creating an Ethical Intelligent Agent. *AI Magazine, 28*, 15-26.

Anderson, R. (2011). Collections 2021: The future of the library collection is not a collection. *Serials, 24*(3), 211-215.

Anthes, G. (2016). Search Engine Agendas. *Communications of the ACM, 59*(4), 19-21.

Baase, S. (2013). *A Gift of Fire: Social, Legal, and Ethical Issues for Computing Technology* (4th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Birkbak, A., & Carlsen, H.B. (2016). The public and its algorithms: comparing and experimenting with calculated publics. In Amoore, L. and Piotukh, V. (eds.), *Algorithmic life* (pp. 21-34). London: Routledge.

Borner, K. (2010). *Atlas of Science: Visualizing what we know*. Cambridge, MA: MIT Press.

Bostrom, N., & Yudkowsky, E. (2011). The Ethics of Artificial Intelligence: In William Ramsey, Keith Frankish (eds.), *Cambridge Handbook of Artificial Intelligence*. Cambridge (UK): Cambridge University Press.

Bozdag, E. (2013). Bias in Algorithmic Filtering and Personalization. *Ethics of Information Technology, 15*, 209.

Brunton, F., & Nissenbaum, H. (2013). Political and Ethical Perspectives on Data Obfuscation. In Hildebrandt, de Vries (eds.), *Privacy, due process and the computational turn*. London: Routledge.

Chin J., & Wong G. (2016). China's New Tool for Social Control: A Credit Rating for Everything. *Wall Street Journal, Nov. 28, 2016*.

Clement, T., Hagenmaier, W., & Knies, J.L. (2013). Toward a Notion of the Archive of the Future: Impressions of Practice by Librarians, Archivists, and Digital Humanities Scholars. *The Library Quarterly, 83*(2), 112-130.

Cox, R.J (2007). Machines in the archives: Technology, and the coming transformation of archival reference. First Monday, 12, 11.

Crawford, A. (2015). *The Meaning of the Library: A Cultural History*. Princeton (UK): Princeton University Press.

de Hert, P., & Lammerant, H. (2016). Predictive Profiling and its Legal Limits: Effectiveness Gone Forever? In Bart van der Sloot; Dennis Broeders; Erik Schrijvers (eds.), *Exploring the Boundaries of Big Data (WRR-verkenning, Vol. 32)* (pp. 145-173). Den Haag: WRR.

Dennett, D.C. (2013). *Intuition Pumps and Other Tools for Thinking*. New York: W. W. Norton & Company.

Diakopoulos, N. (2016). Accountability in Algorithmic Decision Making. *Communications of the ACM, 59*(2), 56-62.

Danielson, E.S. (1989). The ethics of access. *American Archivist. 52*, 52-62.

Domingos, P. (2015). *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. Ingram Publisher Services US.

Eco, U. (1980). *The Name of the Rose*. London: Vintage, Random House Publishing.

Etzioni, A., & Etzioni, O. (2016). Designing AI Systems that Obey Our Laws and Values. Communications of the ACM, 59(9), 29-31.

Ferguson, S., Thornley, C., & Gibb, F. (2016). Beyond codes of ethics: How library and information professionals navigate ethical dilemmas in a complex and dynamic information environment. *International Journal of Information Management, 36*(4), 543-556.

Fernandez, P. (2016). Through the looking glass: envisioning new library technologies. *Library Hi Tech News, 3,* 20-23.

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The Rise of Social Bots. *Communications of the ACM, 59*(7), 96-104.

Floridi, L. (2014). *The 4th revolution: How the infosphere is reshaping human reality* (paperback version). Oxford: Oxford University Press.

Ford, M. (2013). Could Artificial Intelligence Create an Unemployment Crisis? *Communications of the ACM, 56*(7), 37-39.

Garoogian, R. (1991). Librarian/patron confidentiality: an ethical challenge. *Library Trends 40*(2), 216-233.

Gomez-Uribe, C.A., & Hunt, N. (2015). The Netflix Recommender System: Algorithms, Business Value, and Innovation. *ACM Transactions Manage. Inf. Syst. 6*(4), Article 13.

Granka, L.A. (2010). The Politics of Search: A Decade Retrospective. *The Information Society 26*(5).

Herring, M.Y. (2014). *Are libraries obsolete? An argument for relevance in the digital age*. North Carolina: McFarland & Company Inc. Publishers.

Jordan, M.I., & Mitchell, T.M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255-260.

Kallberg, M. (2012). Archivists 2.0: Redefining the archivist's profession in the digital age. *Records Management Journal, 22*(2), 98-115.

Kirchhoff, T., Schweibenz, W., & Sieglerschmidt, J. (2009). Archives, libraries, museums and the spell of ubiquitous knowledge. *Archival Science, 8*(4), 251-266.

Kizza, J.M. (2013). *Ethical and Social Issues in the Information Age*. London: Springer.

Kirkpatrick, K. (2015). The moral challenges of driverless cars. *Communications of the ACM 58*(8), 19-20.

Kosinski, M., Stillwell, D., & Graepel, Y. (2013). Private Traits and Attributes are Predictable from Digital Records of Human Behavior. *Proceedings of the National Academy of Sciences, 110*(15), 5802-5805.

Lancaster, F.W. (1982). *Libraries and Librarians in an Age of Electronics*. Arlington, VA: Information Resources Press.

Laudon, K.C. (1995). Ethical concepts and information technology. *Communications of the ACM, 38*(12), 33-39.

Lichocki, P., Billard, A., & Kahn, P.H. (2011). The ethical landscape of robotics: Bringing ethics into the design and use of robots. *IEEE Robotics and Automation Magazine, 18*(1), 39-50.

Licklider, J.C.R. (1965). *Libraries of the Future*. Cambridge, MA: The MIT Press.

Liu, G. (2011). The Application of Intelligent Agents in Libraries: a Survey. *Program, 45*(1), 78-97.

Mayer-Schönberger, Viktor, & Kenneth Cukier (2013). *Big Data: a revolution that will transform how we live, work and think*. Boston / New York: Houghton Mifflin Harcourt.

McCarthy, J. (2007). *What is artificial intelligence?* (unpublished). Stanford University, retrieved from http://www-formal.stanford.edu/jmc/whatisai/

McCartney, R., & Anderson, M. (1996). *The believability of Road Runner cartoons: logical consistency conquers unreality (AAAI Technical Report WS-96-03)*. Retrieved from https://www.aaai.org/Papers/Workshops/1996/WS-96-03/WS96-03-002.pdf

Medina, E. (2015). Rethinking algorithmic regulation. *Kybernetes, 44*(6/7), 1005-1019.

Metz, C. (2016a). AI Is Transforming Google Search. The Rest of the Web Is Next. *Wired 02.04.16.*

Metz, C. (2016b). The Rise of A.I. What the AI Behind AlphaGo Can Teach Us About Being Human. *Wired 05.19.16.*

Mitchell, T.M. (2009). Mining Our Reality. *Science, 326*(5960), 1644-1645.

Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2).

Moldovan, B., Moreno, P., van Otterlo, M., Santos-Victor, J., & De Raedt, L. (2012). Learning Relational Affordance Models for Robots in Multi-Object Manipulation Tasks. In *2012 IEEE International Conference on Robotics and Automation (ICRA)*, (pp. 4373-4378). https://doi.org/10.1109/ICRA.2012.6225042

Morris, S.L. (2009). An Introduction to Archives for Librarians. *Libraries Research Publications*, Paper 103. Retrieved from http://docs.lib.purdue.edu/lib_research/103

Nilsson, N.J. (2010). *The Quest for Artificial Intelligence*. Cambridge (UK): Cambridge University Press.

Ng, I.C.L., & Wakenshaw, S.Y.L. (2017). The Internet-of-Things: Review and research directions. *International Journal of Research in Marketing, 34*, 3-21.

Noh, Y. (2015). Imagining library 4.0: Creating a model for future libraries. *The journal of academic librarianship, 41*(6), 786-797.

Pariser, E. (2011). *The Filter Bubble: What The Internet Is Hiding From You*. New York: Penguin Press.

Paulus, M.J. Jr. (2011). Reconceptualizing Academic Libraries and Archives in the Digital Age. *Libraries and the Academy, 11*(4), 939-952.

Preisig, A.V., Rösch, H., & Stückelberger, C. (eds.) (2014). *Ethical Dilemmas in the Information Society: How Codes of Ethics Help to Find Ethical Solutions, Papers from the IFLA/FAIFE Satellite Meeting 2014*. Retrieved from https://www.ifla.org/files/assets/faife/publications/misc/ethical-dilemmas-in-the-information-society.pdf

Puschmann, C., & Bozdag, E. (2014). Staking out the unclear ethical terrain of online social experiments. *Internet Policy Review*, 3(4).

Richards, N.M. (2015). *Intellectual Privacy: Rethinking Civil Liberties in the Digital Age*. Oxford: Oxford University Press.

Richards, N.M., & Smart, W.D. (2016). How should the law think about robots? In Calo R., Froomkin A.M. and Kerr I. (eds.), *Robot Law* (pp. 3-22). Cheltenham: Edward Elgar Publishing.

Richardson, K. (2016). Sex Robot Matters: Slavery, the Prostituted, and the Rights of Machines. *IEEE Technology and Society Magazine, 35*(2).

Russell, S.J., & Norvig, P. (2009). *Artificial Intelligence: A Modern Approach* (3rd ed.). Harlow: Pearson Education Ltd.

Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Dziurzynski, L., Ramones, S.M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M.E.P., & Ungar, L.H. (2013). Personality, Gender and Age in The Language of Social Media: The Open-Vocabulary Approach. *PLOS One, 8*(9), 1-15.

Shanahan, M. (2015). *The Technological Singularity*. Cambridge, MA: The MIT Press.

Soares, N. (2015). *The Value Learning Problem - Machine Intelligence Research Institute, Technical report 2015-4*. Retrieved from https://intelligence.org/files/ValueLearningProblem.pdf

Steinert, S. (2014). The Five Robots - A Taxonomy for Roboethics. *International Journal of Social Robots, 6*, 249-260.

Stone, P., Brooks R., Brynjolfson, E., Calo, R., Etzioni, O., Hager, G., Hirschberg, J., Kalyanakrishnan, S., Kamer, E., Kraus, S., Leyton-Brown, K., Parkes, D., Press, W., Saxenian, A., Shah, J., Tambe, M., & Teller, A. (2016). Artificial Intelligence and Life in 2030: One Hundred Year Study on Artificial Intelligence - *Report of the 2015-2016 Study Panel. Stanford, CA: Stanford University*. Retrieved from https://ai100.stanford.edu/sites/default/files/ai100report10032016fnl_singles.pdf

Svensson, M., and Rosengren, C., & Astrom, F. (2016). *Digitalization and Privacy: A systematic literature review*. Lund: Lund University (Media-Tryck).

Tanz, J. (2016). Soon We Won't Program Computers. We'll Train Them Like Dogs. *Wired 05-17-16*.

Tene, O., & Polonetsky, J. (2014). A Theory of Creepy: Technology, Privacy, and Shifting Social Norms. *Yale Journal of Law and Technology, 16*(1).

Thomas, M.A. (2000). Redefining Library Space: Managing the Co-Existence of Books, Computers, and Readers. *Journal of Academic Librarianship, 26*(6), 408-15.

Tutt, A. (2017). An FDA for Algorithms. 69 *Admin. L. Rev, 83*.

van Est, R., & Gerritsen, J. (2017). *Human Rights in the Robot Age: Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality*. Den Haag: Rathenau Insituut.

van der Sloot, B., Broeders, D., & Schrijvers E. (eds.). (2016). *Exploring the Boundaries of Big Data, Netherlands Scientific Council for Government Policy (WRR)*. Retrieved from https://www.ivir.nl/publicaties/download/1764

van Otterlo, M. (2013). A machine learning perspective on profiling. In Hildebrandt and de Vries (eds.), *Privacy, due process and the computational turn* (pp. 41-64). London: Routledge.

van Otterlo, M. (2014a). Automated Experimentation in Walden 3.0.: The Next step in Profiling, Predicting, Control and Surveillance. *Surveillance and Society, 12*(2).

van Otterlo, M. (2014b). *Broadening the Privacy Concept in the Digital Age: Adjusting rights?* (unpublished, position paper prepared for the Amnesty International Strategic Investigations Symposium on surveillance and Human Rights, november 2014.

van Otterlo, M. (2016a). The libraryness of calculative devices. In Amoore, L. and Piotukh, V. (eds.), *Algorithmic life* (pp. 35-54). London: Routledge.

van Otterlo, M. (2016b). Project BLIIPS: Making the Physical Public Library more Intelligent through Artificial Intelligence. *Qualitative and Quantitative Methods in Libraries (QQML), 5*, 287-300.

van Otterlo, M., & Feldberg, F. (2016). Van kaas naar big data: Data Science Alkmaar, het living lab van Noord-Holland noord. *Bestuurskunde, 1*, 29-34.

Vinge, V. (1993). The Coming Technological Singularity: How to Survive in the Post-Human Era. Retrieved from https://www-rohan.sdsu.edu/faculty/vinge/misc/singularity.html

Wasik, B. (2013). In the Programmable World, All Our Objects Will Act as One. In *Wired (05-14-13)*. Retrieved from https://www.wired.com/2013/05/internet-of-things-2/

Wiering, M.A., & van Otterlo, M. (2012). *Reinforcement Learning: State-of-the-Art*. Springer-Verlag.

Wright, A. (2014). *Cataloging the World: Paul Otlet and the Birth of the Information Age*. Oxford: Oxford University Press.

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *PNAS 112*(4), 1036-1040.

Zastrow, J. (2013). Digital Changes Everything: The Intersection of Libraries and Archives. *THE OPINION PIECE, 33*(9).

Zuboff, S. (2015). Big Other: Surveillance Capitalism and the Prospects of an Information Civilization. *Journal of Information Technology, 30*, 75-89.

ARNOUD GLAUDEMANS, RIENK JONKER AND FRANS SMIT

Beyond the traditional bounderies of archival theory: An interview with Eric Ketelaar[1]

EDITORS: Before going into discussing the key issues of the articles in the book, we would like to start with a more general question. We entitled the book 'Archives in liquid times', because we have the impression that on many levels – concepts, fundamentals, ethics – our profession is in so much movement, that the metaphor of being 'liquid' that we derived from sociologist Zygmunt Bauman, is an appropriate one. What is your opinion on the state of affairs in our profession and in archival theory? Would you agree that we live in liquid times?

ERIC KETELAAR: On the one hand I agree that it is a time of different views. You know that I have been influenced by postmodernist thinking. So, I would agree that nothing is stable, and things are always changing. In my view a record is never finished and is always actual. As to the metaphor of 'liquid', for me it sounds slightly negative: as if nothing can be taken for granted, nothing is sure, and nothing is stable. As far as our profession is concerned that is a bit 'too much' and too negative. The fact that nearly all the essays in your book treat basic concepts of archivistics – like context, provenance, etc. – shows that these concepts still are 'places to reside'. They are not so strict and immutable as some people would think, but you can trust those concepts. Each of them has a history. Of course, our profession is undergoing a lot of changes, as can be read in the essays in your book. Also, the fact that you, editors, started the whole exercise of making a book shows that you yourself are not as insecure as Baumann's metaphor suggests!

As I have been arguing for a long time already, our profession as such, and even more than the library profession, tends to be quite on the conservative side. This is what Van der Gouw already stated in his inaugural address about forty years ago. In the case of the Netherlands, the fact that the archival profession was so early professionalised, and that we 'had our bible before our church', caused that for a very long time the profession did not really evolve. Also, recall that Fruin stayed on as National Archivist until his seventies; he controlled the State Archives and indeed the whole profession as President of the Society of Dutch Archivists, as archival educator (he founded the School of Archives (*Archiefschool*) and held the chair of its examination board) and, after 1920, as the only surviving author of the Manual. Together with Muller, he had a very conservative influence and view on the profession. It took quite a while for younger professionals, in the 1980s, to liberate themselves from a one-dimensional view of archivistics. This is also to be coupled with the natural tendency of any professional – in for example medicine, law, etc. – of being cautious in his or her treatment of theory. Recall that in my Leiden and also

---

in my Amsterdam inaugural address I quoted American writers who claimed that archival theory is 'much ado about shelving'. However, I have always argued that theory is an important aspect of our profession, as stated in the first sentence of Glaudemans' and Verburgts essay. And, coming back to the point of 'liquid' times, as an inveterate optimist I intend to take postmodern 'liquidity' in a positive way, as a challenge, and not – like Baumann seems to do – as something to be very sombre about.

EDITORS: So, to you the 'liquidity' has its dangers but also gives a range of possibilities to improve things. We, however, adopted this term because we are convinced that in the archival profession we need more fundamental thinking, and that we are missing it – which was also the reason for this book.

ERIC KETELAAR: As I already said, that is what I appreciate and admire in the book; that it is an endeavour to look beyond the traditional boundaries of archival theory, trying to learn from other disciplines. Still, there are some disciplines missing in your book. For instance, in my chapter in the *Archival Multiverse* I also mention sociology, anthropology, performative and dance studies. One has to avoid thinking that your book covers everything. In this sense it is 'open-ended', as can be also be seen in the fact that some articles – e.g. Van Bussel – point out areas of further research. Do not fall into the same trap our profession has fallen into in the past, by fencing off the boundaries. Unless you believe – wrongly, I think – that it is only data science, or mainly the algorithmisation of the world, we should turn to. There are many more relevant disciplines.

EDITORS: When we started making this book in 2014, our first idea was to get involvement from the field of information philosophy. From that community there were not a lot of people who talked to archivists.

ERIC KETELAAR: What is striking and significant, if you look at the authors of the articles in your book, is that you have a lot of people working in the archival field, and very few really from outside. It shows a more general epistemological problem. In 2012 I wrote a short piece in *Archival Science* about ten years of archival science. The original idea of the journal was, that we would reach out to other disciplines. When five years ago I checked this, I noticed how few references to archival literature I found in non-archival journals. Thus, the interesting question remains: how, as professionals, can we achieve a reciprocal exchange of ideas with other disciplines, and which disciplines should be on the top of our list?

EDITORS: Perhaps we can make a distinction between two types of disciplines: one concerning fundamentals, and structure of information, and one more concerning use and reuse of information. For there is a difference between the two. From this distinction we might be able to connect to other disciplines better.

ERIC KETELAAR: Instead of a book, again destined mainly for the archival community, you could also take the articles and submit them to a journal on psychology, anthropology, etc. – and see what happens. Everyone agrees that we should have a multidisciplinary approach, but it is very difficult to realise. Also, the moment you borrow concepts from other disciplines, you are often accused of gate-

crashing. A future challenge for all of us remains: how can or should we go on
inviting people from other disciplines to reflect on our discipline on the one hand,
and on the other, how can we stimulate people working in the archival field to – at
least – take note of what is happening in other disciplines? You as editors, although
implicitly, make a very strong case for information philosophy as the 'saviour'
discipline. I am not quite convinced of this. As to the concept of information, I agree
with Geoffrey Yeo that archival documents do not contain information, but that
'information is just one of many affordances obtainable from records'. So, I wonder:
what do we have to do with information philosophy? Could one be more specific as
to what the two-way influence might be, here, between archivistics and information
philosophy? Will, for instance, the keyword 'archives' be used a hundred times more
in Floridi's next book as a result of the contacts the archival profession initiated?

EDITORS: It seems that nowadays one of the themes in information philosophy, or
to be more specific, in information ethics, is on archivists as keepers of trustworthy
information, having to be more active towards the upcoming fake-news
phenomenon. The focus seems to be 'past' the function of archives in government,
towards on the 'good' exchange of big data – for instance from multinational car
companies to the medical field.

ERIC KETELAAR: In information philosophy one apparently has a false, or at least
incomplete, idea of what archival science is or could/should be. To me, Archives
(with a capital A) have nothing to do at all with post-truth (the fake-news
phenomenon). The post-truth tweets by president Trump are reliable, genuine
presidential records. That they are falsehood, is simply something else. As I have
often said, the oldest archival document in the Netherlands from the year 1000 is a
fake, a forgery (or, in diplomatic language: a *seeming* original). I do not think it is up
to the archivist to say: I do not want to have the false or fake ones. Then, half of the
material in the Dutch National Archives could be thrown away. It is up to the user –
the politician, the journalist, the historian – who in some years' time studies e.g. the
Trump tweets, to discover whether or not they were a representation of a false or a
true fact. However, I would propose that archival institutions and archivists could
assist in such an assessment, through answering questions like: what was the
context, what was the business process, what was the archival bond with other
documents – for instance between the tweets and records of cabinet meetings. But, I
do not think we have a particular job in evaluating factual truth or falsehood. The
value of a record is in the eye of the beholder, so to speak. What is true for one may
be not true for another person. It comes down to the question of who is responsible
here. Take the records manager: does he or she have to tag the presidential tweets of
'true' or 'untrue'? I do not think so. The tweets being records is not under discussion
here. It is the same with diplomatics: you can vouch for the reliability and
authenticity of the record as such, but whether the information or the message in
that record is true or not true, is not for the records manager to decide.

EDITORS: That is exactly the difference between information philosophy and
archivistics. Like in the Manual of 1898, archivists are responsible for the object, not
for the content. When reading Floridi, you can conclude he does not focus on the
trustworthiness of the object, but on the truthfulness of the content. That is a very
big difference of the two disciplines. The question is how to build a bridge between

the two. On the other hand, there are some conceptions in Floridi's information philosophy that, we think, are very useful for archival reasons. So, we have to build bridges knowing the differences.

An important thing Floridi stressed in our last interview was that nowadays information is very much 'in your face', but that we forget, or have little idea of, the actual materiality of digital information. An example given was the bit-coin: without electricity, these coins would simply disappear. Materiality of the digital is also very much our concern, because it is about preservation.

ERIC KETELAAR: In my keynote speech at the ICA congress in Seoul last year I dealt with the materiality of the digital, invoking the archives of Salman Rushdie, now in the Emory Libraries in Atlanta (Georgia). These archives consist of paper, hard-disks, four PC's and some CD-ROMs. These are materials or objects, with information. You would lose something when you would migrate them to, let's say, one new medium. So, the library in Georgia made a simulation where you can simulate being – like Salman Rushdie – behind a Macintosh Performa 5400 (a computer that was on the market in 1997/1998). I contend that we have to preserve a lot of these materials, these objects, because only then we can render how information or records were used in the first place. I contend that every archive in an archival institution is not the original and authentic archive anymore, because: what do we do? The archives arrive, in a certain order or disorder, and we start with unpacking and repacking them in new folders and in acid-free boxes. I would like to see a simulation of how, let's say, the records of the *Staten-Generaal* at the National Archives in the Netherlands, were used in their primary context (the current display in the National Archives is a laudable effort, but it is not dynamic and not comprehensive). You know Derrida's 'the mutation in technology changes the content of the archive'; the assumption that the sender of an email expects an answer within some seconds, influences what you are writing in your email. In order to keep knowledge of how these records were created and maintained, we would have to preserve or to emulate much of the digital infrastructure.

EDITORS: To go back to Floridi: could you not conclude from what he stated in the last interview, that archives are getting a bit marginal in, let's say, Google-society? Doesn't this show that, within the government, as is also our personal experience, we are losing grip on important information with an archival function; so, that when we keep on going with an 'old school' record-based approach, we might end up not having the relevant information. That might be a reason to go into a more 'native' way of thinking about digital information – data science, data quality, etc. – through which you could do better capturing of the algorithmic functions used in government or in governmental processes. The alternative of 'getting marginal' would certainly be a bad and scary thing, so maybe you should ask other questions.

ERIC KETELAAR: I agree, but it is important to note that I am not interested primarily in the archival objects. I am interested in the archive as a process. When you look at the process, you automatically go to where the archive process starts. Which, in government, is at the desk of the civil servant, or the minister. So, we have to move our focus to the beginning of the process. Take the MH17 airplane-crash in Ukraine in July 2014: 'the MH17 archive', as members of parliament want it to be, will not consist of a big archive, constructed *ex post*. It is rather an intertextual body

of records created and still being created by various record creating agencies. The National Archives cannot sit still and wait until at some point in the future those records will be transferred to the archives: their task is to act proactively to safeguard the creation, maintenance and accessibility of that body of records.

Coming back to the post-truth tweets: archivists might meet Floridi, in a way. Not so much because we are concerned about whether or not the tweet is true, but because we are concerned about the government's accountability through its tweets as well as other records. It is the accountability issue that is the most important point, and for that you need context. The context is all: context is the most important, essentially archival, principle, or concept. That is our strength; and we should make it clear to the information professionals and information philosophers, that in their quest for trustfulness of information they should pay attention to context, and even more: they should adopt our concept of records created within a particular context, with an archival bond. Only then can we, or anyone, trust information. Of course, in the digital age, that is the so-called fluidity, things are changing. The principle of provenance, for instance, as applied in the paper age is still valid, I think, provided you adapt the concept and make it usable in the new environment.

Maybe you know there already are a number of proposals for a so-called 'participatory model' for appraisal, access, etc. I would go one step further, and say it is not so much the user you have to account for, no: the user is part of the process. The record's subjects (for example, a citizen) are still seen by government as the object, or the destination, instead of seeing them as co-creator. Take appraisal: luckily in the Netherlands we involve non-archival experts in our procedures, but I have never heard of, for instance, any student being involved in the establishment of appraisal schedules for universities. In a participatory model, any process starts with an individual, somewhere. I could quote a lot of historical examples here. How did recordkeeping by the cities start? Citizens asking city government to authenticate their transactions (and of course, by making lists of taxpayers and by preserving and copying the city charters).

EDITORS: The authors in this book address concepts, principles, models and ethical issues. Could you reflect on their contributions?

ERIC KETELAAR: In his first contribution, **Van Bussel** delivers a critical overview of archival theory since the nineties, which he denotes as a time of 'archival renaissance' – a term that I like. In general, I find his criticism of the records continuum model a bit too negative; to a large extent I agree with his critique of diplomatics. I believe Van Bussel is right in assessing that, in the archival renaissance of the nineties, the question of what is archive, and how the archive is created, maintained and managed, got less attention than it should have had. This is discussed in Van Bussel's second paper.

Ernst's first paper I found difficult, but it provides a very important acquaintance with media archaeology. The second on audiovisual media is a bit more practical. Archivistics could learn a number of methodologies from media archaeology, but we have to keep in mind that we are dealing with two different disciplines here. To a certain extent both disciplines meet in the materiality – as discussed before – of

objects, of infrastructure: in taking the media as such as the object of research, rather than their informational content. Ernst's approach could be useful given the fact that we are moving into a world of less text and more moving images.

EDITORS: Would you also say that the materiality of what you want to capture can be seen as context: the medial possibilities, impossibilities and consequences as Ernst sketches, for instance of audiovisual media, who work differently than text?

ERIC KETELAAR: No, context is a concept. Whether context is a material manifestation or not, does not make any difference for the context. In both cases, material (object) and immaterial (process-bound), you will have to capture the context(s) in which records have been created and used. Take installation art, and see how artists deal with the archive. There are many installation artworks where the viewer, user or any individual is part of the installation. The moment he or she walks into the room, the installation 'starts'; a video or something. What is material here, and what not? I would say that the materiality of the installation has to be captured somewhere. Even for an immaterial performance. Or if we take an example I use in my foreword of ballet, of dance; it is impossible to capture the whole thing, so you will have to rely on capturing as much of the materiality as possible, and knowing that it will never be complete. If only because even a video of a ballet on stage does not take into account the fact that there also are viewers taking part in the performance. Why do we still perform Shakespeare's plays? Because every play – every instance – is different from the other; and even in one season, the interaction of viewers and performers differs.

The ubiquity of data as sketched by **Jeurgens** will have a lot of consequences for record creation and capture. The same goes for the cooperation between human and machine. Other very important issues raised in the article are the 'mechanisation' of record creation, and appraisal, not so much for cultural heritage, but as a means to enhance the accountability of the archive or record creator. Jeurgens stresses rightly that we have to 'revisit' or review what appraisal can and should be when confronted with today's ubiquity of data. You should, somehow, make a distinction between 'raw' data and records in the legal sense.

As to the contribution by **Glaudemans** and **Verburgt** I think, as discussed before, it is not right to prioritise an object-like concept of a record; it should not be taken as a product but as a process. I have doubts about the claim that in the digital, two domains currently exist where consignation takes place: for the record has to exist somewhere, even in the cloud it exists, so the cloud is a place of consignation just as the national archives are. Also, consignation is not so much about fixity, I understand it as a much broader term. As to Flusser, whose work I did not know before, I agree that the fact that we are sender and receiver, at the same time in the same space, will have enormous consequences. I do not expect though, that this implies the disappearance of the distinction of public and private (as Flusser stresses); not in all cases. It depends what Flusser means with public; if you and I are sending and receiving together, I would still regard it as private.

I am struggling a bit with **Yeo** writing that a record does not contain information, but that information is one of the 'affordances' – next to for example evidence, accountability, identity – 'that arises from engagement with records'. So, he makes a

distinction; you have a record, but that record only contains information to the extent that there is someone who is using that record. That must be a very sad message to the information philosophers: saying that a record does not contain information! In a sense he is right, in that an object has no intrinsic meaning in itself. The meaning is given by the one who is interacting – viewing, using, etc. – with the object.

EDITORS: So, the definition of information as meaningful data would fail?

ERIC KETELAAR: Well, it does not fail, but the question is: to whom and when is data meaningful? If I understand Yeo rightly, he stresses that it becomes meaningful only when someone attaches meaning. This someone, one might add, could even be a machine. Would this be true?

EDITORS: We cannot talk for Yeo, but when you 'push it' like this, he might not hold on to this view. Maybe he is saying: let's put the informational aspect aside, and focus on the performative aspects of a record, maybe even an algorithm.

ERIC KETELAAR: Let's take an author who is writing on a piece of paper; that writing has a particular meaning for him or her: the author. The moment he or she pushes the button 'save' or 'send', he or she is no longer 'meaningful' and can't influence the receiver in sharing the authorial meaning. The receiver may attach a totally different meaning. So, the data are there, and in the context of creation they had a particular meaning, but the moment they are transferred along the chain to another person or machine, it is not assured that the meaning stays the same. As I said earlier: the value of a record is in the eye of the beholder. It may change over time, and in different spaces. Each activation of the record adds meaning, new meaning, that may be different from the meaning of the author. So, I understand Yeo when he says: records do not contain information, they make it possible. I am certain that Yeo's account of speech acts gives us an important additional methodology to look beyond the object of the record and focus on how that record has been created and used; to view records as a particular form of social practice in a cultural context.

I agree with **Jonker**'s statement that 'An archivist can only ensure and ascertain that an information object under his control has the same quality as when it was ingested, it is trustworthy with guarantees about the integrity. An archivist cannot and from an ethical viewpoint may not deliver a statement about truthfulness of the information.' But I would go one step further, and say that the archivist can assist or help the interpreter by providing, from his discipline, data about the context. Take for instance the example of someone in the reading room of an archival institution who comes up to the archivist on duty and asks: I have a record here and do you think, is it true or not? Then the archivist should answer: that is not my business, but I can see that the paper has been tampered with; I can see that this document was bound later in the file, etc.

Jonker writes: 'We as archivists are in this fluid digital world looking for fixation points; we want to be able to fixate moments. But to be able to fix, it must be clear why something has to be fixed, what content is to be fixed and eventually how this fixation is to be carried out technically.' Is the characteristic of a fluid world not that you cannot and should not fix things?

301

EDITORS: Yes, it is a contradiction. But at some moment you need to fixate information; when it ends up in a document, a register etc. Make it persistent through time, in some way. This is one of the problems our liquid (or fluid) times bring us. When you do not do that in some way, then we have no archive at all.

ERIC KETELAAR: But does this not contradict the whole idea of liquid times? Do you envision that, although we are in liquid times, we still need persistent representations?

EDITORS: Yes! For accountability purposes, for instance. And in a hundred years from now you would want to look back; that is also what an archive is for. It is that simple.

ERIC KETELAAR: In his second contribution, **Van Bussel** stresses that existing models and theories within archivistics have to be complemented, at least by theories from organisation science and information science; and he is focusing on the organisational side. What I like especially is the attention given to the behavioural, cultural aspects of record creation and use. Recordkeeping is not a mere technical process, but there are people involved. Van Bussel is right in stating that my concept of archivalisation has been referred to a lot, but that is has not yet been really tested in a practical situation. I tried to do that in my comparative studies. There is a practical link of this organisational model and the information model of Jonker. This is something that could be addressed further in discussions, for example in a reading group.

The essay of **Foscarini** and **Ilerbaig** also focuses on the organisational aspects of recordkeeping. The authors agree with Heather MacNeill that diplomatics should be situated within the framework of other – philosophical, disciplinary – perspectives (a view which, by the way, seems not to be accepted by Duranti). Records are taken as communicative events, as forms of social practice, and these are approached through the lens of the structuring function of genres. This is a more 'bottom-up' approach, compared to diplomatics, which is forcing a set of requirements 'top down'. In my view of social and cultural archivistics I rather prefer the 'bottom-up' approach: to really study how people are creating and using records in practice. Archivists, as Foscarini and Ilerbaig rightly stress, should not lose contact with the daily practice of creating and using records.

EDITORS: A current (Dutch) example of such a practice would be the 'archiving by design' approach: to 'architecturally' build in the archival functionality in applications used for controlling the environment. These applications are to be used in cooperative networks executing governmental tasks and processes by a large number of governmental organisations and other stakeholders. Here you will also need 'bottom-up' knowledge, otherwise it will never work.

ERIC KETELAAR: It also has something to do with what Wanda Orlikowski called the 'duality of technology'; a particular technology shapes the way people are behaving, but the behaviour also shapes technology. One of the most famous examples is text-messaging; it was designed as a means to link the customer to the provider, but then some users discovered it could also be used to communicate with

others. As to archival (re-)use, you could or should observe and study, for instance, how people who are actually using search facilities are shaping those facilities and are being shaped (for example: restricting their research to digital records).

I have doubts about **Smit**'s use of Timothy Morton's concept of a hyperobject. Is it meant as a metaphor or as reality? Should I take 'hyperobject' as the totality of all records, or of all digital information?

EDITORS: The term should be taken as denoting a phenomenon that is so vast that it is difficult or impossible to oversee it and get a grip on it. For instance, Dutch philosopher René ten Bos uses the term to describe bureaucracy. In this vein it denotes our digital environment, a bit metaphorically, and comparable to how archaeologists consider the soil as the '*bodemarchief*' (soil-archive).

ERIC KETELAAR: So, 'hyperobject' denotes the whole of digital data, of which records are a part. Could you not argue then, that because it is a hyperobject, the solution lies in decomposition through focusing on the records part of it? That would 'destroy' or make superfluous the concept of hyperobject, because records would remain approachable and overseeable.

EDITORS: In daily practice we are confronted with a way of using and reusing data and records, in which the distinction between the two is blurring to a large extent. For example: how much 'recordness' remains in information when being reused in an open data context? The general problem here is that one, in fact, cannot easily identify the record. This is not only an operational problem, because capturing a record is also a matter of conceptualising. Through analysing, for instance, the aspects of context in the digital environment, you could define how you capture.

ERIC KETELAAR: I would agree, for this discussion, that a record becomes a record only when it is captured in a recordkeeping system. The problem lies in the ubiquity of data. What and when data becomes a record, is a policy decision. I once visited a consulting company, whose main assets were their PowerPoints. I told them: forget all definitions of records and from now on capture these PowerPoints into a recordkeeping system. So, what defines 'record' here is the policy; it is not a question of (archival) theory. I like the idea of the archaeological '*bodemarchief*' but I do not think we need the concept of hyperobject here. As to the arguments about authenticity, and a new way of configuring it – authenticator, authenticating and the informed – : is this a challenge for the archivist?

EDITORS: In our daily work it is. For instance, through stressing the use and implementation of a metadata scheme for records. This, among other things, safeguards the necessary authenticity.

ERIC KETELAAR: So, you capture records in a recordkeeping system, then it is moved to, let's say, the recordkeeping system of an archival institution: and throughout its life there should be a continuous authentication. One could define, in the continuum, certain points at which you have to act – a bit like a time stamp. All this ties in with continuum thinking: records are changing over time, and in retrospect one should be able to follow the chain back to 'version zero', through the metadata.

Michetti rightly points out that the principle of provenance is crucial, also in the digital context, in preservation, use, access, appraisal, arrangement, and description. That is important, because mostly provenance is only taken into account when arranging archives. Clearly it is also crucial in other fields and processes. The more so when it is linked to the concept of context.

Van Otterlo explores the ethics of digital archives with an emphasis on the role of algorithms. He describes a shift from 'the intended archivist' towards 'intentional algorithmic versions' that could be linked to other essays. What I do not like is his stressing the gate-keeping role of archivists. As a view of what archivists do, it is an incomplete one, to say the least.

EDITORS: What is your opinion about the idea Van Otterlo's seems to offer, that you could operationalise a code of ethics in a ruling for designing algorithms? This idea of 'automating ethics' could be relevant not only for access, but also for (other) processes of recordkeeping. What would you say about this idea of 'automating ethics'?

ERIC KETELAAR: I would argue that both an ethical position, and an automated facility regulating access, are politically framed. When, for instance, I subscribe ICA's Universal declaration on access, it is a political decision. There is no natural law stating that archives should be accessible to anyone. That is something that is only 200 years old. What most archivists do not realise is that availability, accountability, findability, etc., are not universal and natural laws or principles. It is the law, yes, but the law is only an expression of what society at a particular point in time believes to be right or wrong. Different ethical positions are possible, and choosing one is a political decision. The outcome is therefore constructed, based on a particular world view. The same counts when I automate access, when I design a digital decision-making process to regulate access. I am talking about 'politics' with a small 'p', of course. This is the nice thing about fluidity, about postmodernism. Postmodernism, as Lyotard says, is: not believing the grand narratives, and asking this sort of questions: why is availability necessary? Has that been put to a test? I see a congruence of the political constructedness of, on the one hand the human, ethical decisions about access, and on the other hand, any automated system providing access. The algorithmic procedures have to be designed, and anything that is designed, is designed within a political, ethical, etc. framework. And of course, there is the fear of human behaviour being guided through 'hidden' algorithmic procedures. But you should not forget that human decisions are, in a way, hidden as well.

EDITORS: Algorithms seem to have something mystical about them: they are hidden, so you have to fear them. Could we conclude that you would say that the bias in algorithms – hidden or non-hidden – in the creation of information is an aspect of context and not of the 'record itself'? How would you 'situate' this?

ERIC KETELAAR: Well, coming back to what we discussed earlier: if indeed the (informational) meaning of a record is not in the record itself, but in the eye of the beholder or user, you could conclude that in an automated system or algorithmic procedure, the meaning of a record is created by 'the eye' of the algorithm. It would

be a matter of accountability whether we can oppose such a meaning as given by the algorithm; can we bring the algorithm to court, so to speak? Who is ultimately accountable? Accountability is the essential thing here.

EDITORS: When interacting with a human being there is something the Germans call '*freies Ermessen*', which is not possible with algorithms. With a human being you can interact in the sense of interference, discussion, and influence.

ERIC KETELAAR: Algorithms are programmed, and usually they are not 'fuzzy'. Compare also the discussion about robotics and artificial intelligence.

EDITORS: To conclude this interview we would like to ask a last question. Given our doubts and uncertainties concerning the archival theory and profession today: Where, do you think, the profession will stand in, let's say, fifteen or twenty years?

ERIC KETELAAR: I anticipate the development going steady but more slowly than most people think, are afraid of, or hope. The big challenge is that the show must go on, the shop is open, and business has to be continued, while at the same time we have to adjust theory, methodology and practice, and think about what these other disciplines have to tell us. I am concerned that archival professionals will not take time for really sitting down (or standing up) and have the sort of discussions we are having right now. It is important that, in the Netherlands, we continue the professionalisation in the way we started when the training of archivists was moved from the School of Archives (*Archiefschool*) to the university. We now have the opportunity of a more theoretical treatment of archivistics than we had in the past. The same is true for other countries. We should be more open and hospitable to other disciplines, as for example in the United States, with its Archival Education and Research Initiative (AERI). In the Netherlands, as things have developed until now, we still tend to be very much 'inward looking' into our own discipline. We have to be more open to other disciplines. That is pretty much the case already at the university, but it would also be important for practicing archivists and their professional bodies. They should take the time and the opportunity to partake in theoretical and methodological discussions like the one we are having right now. It is of great importance that theory and practice stay closely connected, and we should find structures for that, like reading groups, panels, blogs etc.

I believe that in the next fifteen years archival theory will not change drastically, but will adapt itself to connect with changes in society. I partly base this, my optimism, on the fact that in your book basic concepts of archivistics are being discussed and adapted to a new environment. So, in a way, this collection of essays is a proof that archivistics can move through time: adapting, inhaling, and infusing from other disciplines. Therefore, let us try to keep up and maintain the triangle of theory, methodology and practice. All these aspects should stay closely connected. I am also curious what the increase in foreign students studying archivistics at the university in the Netherlands will bring us, enriching archivistics in this country, and abroad.

So, there are a lot of signs of archivistics being, and staying, a truly vibrant discipline.

ARNOUD GLAUDEMANS, RIENK JONKER AND FRANS SMIT

Documents, Archives and Hyperhistorical Societies: An Interview with Luciano Floridi[1]

EDITORS: Our book, the subject of this interview, contains several essays and/or articles, mostly written by people from the archival community, some by people from elsewhere. We named the book *Archives in Liquid Times*. It is a metaphor from Umberto Eco, who got it from Zygmunt Bauman, a sociologist. This came from the idea or feeling, that today we cannot and should not treat archives from a single paradigm. Our times are way too dynamic for that. The idea of the book is to reflect this dynamic state, and to give a sort of 'snapshot' of our situation. As a preparation we sent you five questions with some clues about the topics we would like to address in this interview.[2]

FLORIDI: The project of your book is very interesting. I remember our past conversation.[3] It is clearly a challenging time, not only for archives. We all know that the analogue world is becoming increasingly digital. I can see that we (you editors and I) belong to the same generation. We experienced the world when it was still entirely analogue, the digital was not present yet. Then, in a matter of just a few

---

[1] The interview was held on September 6th, 2017 at the Oxford Internet Institute (OII).

[2] The interview did not exactly follow the sequence of these questions. The questions were:

1. A proper, contemporary definition of records is given by Geoffrey Yeo in 2009: (…) *it now seems appropriate to characterize records as persistent representations of activities or other occurrents, created by participants or observers of those occurrents or by their proxies; or sets of such representations representing particular occurrents*. The elements of representing and keeping persistency in this definition points to a functional approach instead of an approach where records are seen as material objects. Should the record-based approach to the archive be replaced by this functional approach in terms of algorithmic processing? Obviously, all digital information, including digital records, is processed algorithmically. Could the impact on the archival function of the intrinsic, content-technical, effects of digital mediality (without analog counterpart) be better accounted for from the functional approach?

2. As to data science: Can existing archival concepts, and the specific detailed archival knowledge they entail, be (re-)used in the approach of algorithmic processing, to enhance and enrich the desired functionality in this particular field? And which particular concepts would be relevant in this respect? One could think of concepts like authenticity, provenance, and accessibility.

3. As to cultural criticism ('Kulturkritik'): Which function(s) should the archive and archival theory fulfil today? It seems that historico-cultural criticism, in which the archive functions as a source, is becoming increasingly obsolete and even counterproductive in a more and more post-historical society. Should cultural criticism not be complemented by an account of the archival function within a critique of the infosphere?

4. As to governmental accountability and control: Should the effects of the current and progressing transition to the digital not be accounted for better, in order to avoid an increasingly diffuse allocation of responsibility and distribution and execution of power? A lacking account of the impact of digitalisation might thus have undesirable effects on democratic accountability. What are your views on this matter?

5. As to ethics: In what areas – as addressed in questions one to four – can ethics play a particularly vital and guiding role? Especially, how do 'personal' or 'societal' ethics relate to (an account of) the institutional changes as caused by the digital?

[3] For the 2014 interview see: Glaudemans, A., Jonker R., Smit F. (2015).

decades, we found ourselves living in a world that is completely mixed, with analogue and digital features intertwined. And within this mix, the digital is leading. It takes 'two to tango', but the leading dancer is the digital, at least most of the time. In that sense, the topic of your book and our discussion now, the area of archival studies, is part of 'the big book of revolution'. It is part of the huge transformation we are undergoing.

Recently, in a different context, I was discussing similar issues within the banking system. It might seem a bit far-fetched to compare archives to banks. But when I got your message, it did ring a huge bell. I thought, just imagine how many branches of any bank are closing down for digital reasons. Who needs to go to the bank to do business? Most of the time we are doing everything online. If you need cash, you get it from an ATM. If you need to transfer money or receive a payment, you can do everything online. You do not need to go to a bank anymore. The bank was a place where your *location* physically, and your *presence* interactively, were the same thing. This whole is now split. My interactive presence is now online. My physical, geographical location is no longer necessary. This split has caused the closing down of more than a thousand branches of different banks in this country (the United Kingdom) over the past five years. As to our public libraries, we have the same problem. Location and presence are now *split*. The public libraries are becoming useless if they are only warehouses where to store and get analogue documents. What if I get the documents digitally online, what if I buy the book on Amazon for a fraction of the price that it would cost me to go to the library? To me it seems that the archives are also part of this huge challenge that you, as editors, denote as 'liquid'.

I would like to answer some of your questions concerning the state of affairs.

EDITORS: The first question concerns the nature of records. There are quite some essays in this book which use the word 'materiality' of records. Of course, as old fashioned, but not very traditional people, we know how the record is really the object in the paper world. What we think that is happening, is that – fortunately – many scholars in archival science are drifting away from the idea of materiality, and into the digital world – which is a good thing. The nature of records tends to be defined much more in terms of functionality of information. When we take the definition of records from Geoffrey Yeo as persistent representations (Yeo, 2008, p. 136), we could analyse and understand them in a functional way. This would imply that, in a digital world, a record should not be understood as an object anymore but as functionality through algorithmic processing. Maybe you have some thoughts about this way of thinking?

FLORIDI: Out of your five questions, I found this the most difficult one to answer. So, let me start with a couple of clarifications.

The first is, that there is a misconception about the non-materiality of the digital. It is a dangerous misconception that we are still endorsing these days. It is dangerous because we are relying so much, and increasingly so, on the digital to record our present and future. It is crucial not to forget that the digital resides somewhere. It is usable according to a particular technology and subject to an enormous amount of

risks, like a virus or a wrong click of the mouse. The materiality of digital
information is not immediately clear and obvious, as is the case with printed
information. The materiality of the digital is more remote and hence less visible. So
it is not immediately clear how to take care of the materiality, the physicality of the
digital. Any big company that moves to the digital knows exactly the problems that
arise here. At the Bodleian Library we had this case of someone famous donating his
entire collection to the library – including all the floppies. To read them you need to
buy some old piece of hardware. There is nothing to put the floppies in anymore and
the software is not available. I know that I am telling you things that you know by
heart. But I think what is happening in our culture – generally speaking – is the
following. On the one hand, there is the obvious physicality, the materiality, of a
piece of paper. We know how to take care of it. The material nature of the digital, on
the other hand, is more remote, and therefore we feel less involved. It is also more
difficult to explain that we need to invest resources (financial but also intellectual
and human) to take care of the materiality of the digital. We have to be very careful
about the distinction of digital 'nonmaterial' and analogue 'material'. It is not
adequate, because the analogue and the digital do overlap in many ways.

Let's move to the second clarification, which concerns the distinction between
functional use versus *material identity*. When I first read your question, I thought:
the archive is one of the places of memory, like the museum, the library, or the
gallery. These are places where we accumulate and curate memory. The question is:
what for? Is it just to enable different forms of functionality, like going to an art
gallery to enjoy it, or consulting some documents for a research? Or is it also because
we think memory must be collected in order to protect, preserve and foster our
identities and give sense to our lives; otherwise we would not be who we are and
could not interpret our existence constructively? *Continuity* is what counts here,
because it makes construction of our identity possible. The digital tends to stress
functionality and hence usability more than identity construction and hence the
continuity and meaningfulness of the narratives. Memory is the basis of identity.
The digital is much more about functionality, purpose, usefulness, accessibility,
availability, and so on. The digital again pushes us in that direction, making us forget
that artefacts are the historical memories of who we are and can be. Archives are full
of beautiful things that are not just usable but also fruitful, insofar as they can make
us grow in our self- and mutual understanding.

Given the previous two distinctions, I am not impressed by the definition of 'record'
that you give, as 'persistent representation'. It is not that I disagree, or think it is a
bad definition. I just think it is restrictive, in terms of what archival items are in
their varieties. People who have been to an archive or work with old documents will
know that, for instance, a page was read many times, because the corner of the page
is totally black. The materiality of the thing is consumed by people as well. The
digital does not get licked. It does not fade. If you look at some manuscripts – years
ago I did quite a lot of research on medieval and renaissance manuscripts[4] – there is
a lot of information in the specific materiality that can easily get lost by digitising
everything. When a document is digitised we of course have the picture, the high
resolution of the reproduction. But the digital version comes short in not providing
this kind of materiality. When defining records as a persistent representation in a

---

[4] See Floridi, L. (2002).

functional way, this tends to be forgotten, and that would be a pity. We run the risk that memory may be memorialised (mummified in some permanent, immutable form) by the digital, instead of being kept alive as a trace of the past.

Then there is the whole debate about authenticity, which is of course complicated. The authenticity of the digital has gone through a variety of technical solutions. With the new and ongoing development of solutions like blockchain, the digital seems to regain some credibility in terms of authenticity. In general, though, the digital kills authenticity: there is no sense in asking what the authentic digital document is and what an exact duplicate of it. A file and its copy are identical. There might be a timestamp, but the timestamp is just metainformation, an addition to the file. The two files themselves are interchangeable. They pass the Leibniz test, so to speak: they are not distinguishable, and therefore they are interchangeable. Leibniz used to talk about the identity of indiscernibles. If you cannot discern the difference between A and B, A and B are the same thing. We are used to that, since we started producing things in a repetitive way, from identical vases to identical books. The digital has exacerbated a problem already caused very visibly by the industrial. The more you produce indistinguishable things, like two instances of the same iPhone, the more you lose authenticity as a concept that applies to singular objects (call them tokens) instead of their model (call that type). In this vein, the impact of the digital on authenticity is profound. However, with new technologies, we are trying to find a solution 'with the digital for the digital'. Blockchain is a solution, and there may be others. You can have a series of records that keeps track of a particular file, giving a full history of who originated it, when it was modified, by whom, and how.

This brings us back to the materiality of the digital. Because that blockchain 'lives' somewhere. It lives off electricity, on computers that need energy. That is not trivial. For instance, bitcoin is one of the currencies that use blockchain. And if massively adopted, it would be an environmental disaster. To simplify, it would be like having every coin in your pocket, every banknote you have in your wallet, behaving like an electric bulb that needs to be kept switched on. Clearly, if we were to implement bitcoin as a currency (like the Euro), this would consume a massive amount of energy. That is the materiality we are talking about.

Now let us go back to the distinction I introduced before. The materiality of the digital is of a different type than the materiality of the analogue. You cannot just say that the archives are losing materiality through digitalisation. We are talking about a different kind of materiality. So, we should better discuss the differences of the materiality of the analogue world and the materiality of the digital world. Because there is a difference. We need to figure out what care both kinds of materiality need. To make a difference between material versus immaterial or nonmaterial, is highly misleading.

EDITORS: There is also another aspect of persistence: that of the informational content. Is it still useful to make it persistent? Information nowadays becomes more fluid in how it works, functions, has meaning and is used. Should the term 'record' be interpreted as a function to retain the persistency of the information contained in it?

FLORIDI: Yes, that is a very good point. So, the persistence is not what qualifies an archive. Of course, it has to be there, you have to go back to it, it has to be the same. You might have grasped that my general strategy here is reshuffling existing distinctions rather than endorsing them or abandoning them. In the case of the digital, we have the problem of the fluidity. The digital rewrites itself and is easily modified. How do we cope with the persistence of it? In the best scenario, persistency is kept when a record is going through a series of modifications. Consider for example Google docs. When you write a document, it keeps time stamped copies, so that you can go back to previous versions. That document is not persistent in the same way as a printed piece of paper is, it is malleable, and yet it has a history of continuous changes that we may be able to reconstruct, contrary to the single version of a paper document. This means that the digital can provide a longitudinality of memory (all the several copies of a manuscript, for example) much more easily than the analogue. This is very different from the recent past, when all you could do was rewrite your file every time you saved it. The digital is in a way keeping its 'persistence' by making sure that it keeps a record of the sedimentation of versions. These are like snapshots of its development. Unfortunately, most of the digital information that we have does not enjoy that persistence, like for instance most websites.

The question is: what is the value of persistence? To me, it goes back to authenticity, being able to say: that is what we agreed upon. You could say that it is truthfulness that we are preserving in a document or in a record. Remember that writing was not invented for postcards, or to tell people about our holidays. It was for legal reasons. It was for keeping reliable answers to questions like: how many sheep do I have? how many cows do I owe you? whose land is this? which rules have we agreed upon? I think that that legal ability of records makes people coordinate their actions. It seems to me that the value of persistency goes back to authenticity and therefore to providing the reliability that something was "like this".

In the digital environment, we need to bind the record with the technologies required to read it. Analogue records are not very demanding in terms of technology. As long as there is a bit of light, as long as you can see, as long as you can read, then it is okay. Today that is clearly not the case. A digital record without the right soft- and hardware is as good as a magnetic pattern. In this sense, we are making the materiality of records an issue that is even more serious than in the past. Because now, not only do we have to take care of the material record, but we also have to take care of the software and the hardware required to make it accessible. The problem is so much present that it tends to disappear. It is difficult to explain to people because for us materiality is what you bump into. But we also have the materiality supporting the digital.

EDITORS: You could say that materiality has a lot of layers, which we have to take into account. We account for the necessary layers and the rest we trust until otherwise proven. It is all about trust, not truth. There was always a kind of trust in the paper world. In the digital world it could work the same way. With some processes it is necessary that there is for example a third party to authenticate, and thus provide in this trust.

FLORIDI: I think that is a good way of putting it. There are cultures or societies where there is so little trust that you would want to check everything. I think that there are these flexibilities within our societies about where the trust stops. We really are on the verge of the transformation of the digital, where the digital is becoming more trustworthy by blockchain and the versioning of files. I hope we will always have someone somewhere to certify and authenticate. We need a kind of authority that generates the trust. This is an argument for the authorial sources of trust, like archives, museums, libraries, or galleries, and the ability to generate trust, which would make our digital world much better.

There is one point that connects trust to persistent representation. When you list these values of trust, authenticity, persistence and truthfulness, you can see that, morally speaking, they are not necessarily good in themselves. I can have, for instance, an authentic, legal decision where Jesus is condemned to be crucified. It is not morally good, but it is very authentic, truthful and trustworthy. I think our society is now at a certain point of maturity of thinking about ethics. We finally realise that a lot of things we call morally good are not necessarily morally good in themselves, but they help the moral good to develop. They are conditions that facilitate the morally good. I like to call this *infraethics*. It is the infrastructure of practices that facilitates morally good behaviour. The crucial role of records in our society is conditional and infrastructural. They enable and empower a better society.

Archives and records are maybe more useful than ever, given the massive communication on social media. All that 'liquidity' is not necessarily good. You need stable records to check the value. I disagree with the view of the library and the archive as a warehouse, as essentially a place where you store documents. That was never the real nature of either the library or the archive. They have, and had, a social and political function. Thinking of archives and libraries as warehouses would mean their end. Just as it happened with those bank branches I referred to earlier. We do not need a warehouse. Who goes to the warehouse these days? The view that we are just collecting things like butterflies to put them on shelves or inside drawers is wrong. We will pay for this mistake politically through the gradual corruption and the pollution of the space of information. And this will happen precisely because, in a variety of ways, everybody involved – the archival and library sources included – did not play their role properly.

We should hear the archival people in this country (UK) say 'this is rubbish, this is not true, we have the records if you want to check, you are welcome any day, here is all the good information you need'. There is a political or socially committed role for the archivist to play. The preservation of memory is only half the task, the other half is to ensure that the preserved memory plays a fruitful role in society, reminding us who we are, and what we may be able to achieve collectively.

Of course, there is always a tension here. But archives involve more than just storing and giving information. They should help form and inform the social debate. If we reduce the role of archives and libraries, the places of memory, to warehouses, then we stop talking to society, and this can have very negative effects.

EDITORS: You think it has to be an active role and not a passive one?

FLORIDI: An active role, absolutely. And it is time because society is not going in the right direction. When you see that millions of French people are voting for a fascist party, it is clear that they have forgotten history. Their archives are not talking. And they are silent because they are warehouses. And if you do not go there, they are not telling you anything. And that is a scandal.

EDITORS: We could analyse this in terms of the distinction between *strong* and *weak* authenticity. Strong authenticity is as we discussed earlier: a thing is what it 'purports to be', as the archivists call it. Next to that you have weak authenticity in which for example records are used to construct identities. There the political aspect comes in. For example, the Amsterdam Museum, formerly the Amsterdam Historical Museum, has a city marketing website. The Amsterdam Museum contributes to deliberate citybranding by saying: we are the city of tolerance, the city of Spinoza and the city of higher arts. And in one little sentence they say: oh yes, we cannot avoid talking about it (although we would like to) but the city was also guilty of slavery offences. The curators should play an active role here and say: everything we curate is strong authentic material. Irrespective of any political or marketing argument our collection should be accessible without any restrictions and without any framing based on political or economical bias.

FLORIDI: Yes, the spin here, the story-telling, the framing, the selection of 'forgets', the edits, the undertones... all this is, by the end of the day, the way history is treated, and it is a scandal. Consider the UK. The way we understand British history is not realistic. Many have this deluded view about colonialism for example, where the British are the only ones who actually have a "good colonial past". If you also read other sources, say Indian reports about what colonial Britain was like in India, about the massacres, the wars, the killing, the expropriation, the violence, the arbitrary borders. It is not that all this is denied, but it is never highlighted.

Records are always incomplete, but not all incomplete records are born equal, so to speak. What do you celebrate during the year? One does not apologise once a year for the massacre of some people in some distant colony, of course not. What one celebrates may be the winning of a war, or the day of the declaration of independence. This is what creates social memories and social identity and cohesion. But it is also dangerous. If we 'edit' our memory too easily, like the digital very much enables us to do, we end up in a filter bubble. The people of strong authenticity should act. But it is a big call for a profession that has been a little bit shy and less prone to be in the limelight.

EDITORS: Our next question is more about data science. Are archival concepts, like authenticity and provenance, be relevant in the data science setting? We observe they are not used very often in practice. Or should we assume that these concepts are already part of the functions that are developed in data science?

FLORIDI: No, I think the question you are asking is open. As you know I chair the Data Ethics Group of the Alan Turing Institute, which is the British institution for data science. One of the things we are talking about more broadly is *information quality*. Authenticity is one of these qualities. But reliability, timeliness, accessibility and availability are important as well. One debate we are having at the moment is

about data science being applied and used to obtain information from huge quantity of data, no matter what the quality of the data is. Sentiment analysis of tweets is a good example for analysing how people react to news, for example. When there is an election people tweet a lot of details. Then you have literally millions of messages to analyse. It is a really difficult and slippery job to have a massive algorithmic analysis of data in terms of data science. How can you take all the data and squeeze some good information out of it? This question comes up all the time, especially when you deal with huge databases which have not been curated. Another strategy, which you see taking place in some corners – especially in medical research – is having access to highly curated, high quality small datasets. A typical example here is Google working with health organisations in England, with access to medical records that are way more reliable, truthful and authentic. Here you do not need a million records, but maybe a thousand, as long as they are very good. You must be able to trust them. So, there are two strategies: take huge quantities of data, throw lots of statistics at them and try to squeeze something good out of them, or take smaller, very highly curated sets, and work very precisely on the sort of training of algorithms and useful information you wish to obtain. That is where data science is now exercising different levels of influence.

Now, when it comes to the archival world, you normally find highly curated documents there. That is why the great companies of the world are so interested. Archival material combines two important features: high quantity and high quality. Remember that data science is about using the data – and in this case to train algorithms on them – to get the kind of information you want. Once the training is done you do not need the data anymore. For instance, the machine needs to see ten thousand pictures of cats. Once the machine knows how to recognise a cat, the pictures are not needed anymore. So, when I have many radiographies of a particular kind of cancer, the machine will learn that it is cancer. Once the training is complete, there is no longer need for massive quantities of data. So, in that sense the archival material is a training ground for data science, and it is very precious. All the effort that has been put into providing high quality material is exploited to provide good input or for training the algorithms. The point here is that all the work that has been put into it should be paid.

EDITORS: A lot of those data are in possession of governments. It is free in the sense of open data. So, the government cannot ask money for the data they are delivering.

FLORIDI: This is something I actually had a discussion about in the past. The opening of national archives to free public use should be the norm. However, when the free access to public archives generates income for companies, we might start having a so called *freemium* solution, where people start paying increasingly for how much they actually are exploiting the particular archive, to the point where it is at a full price. Take for example the huge archive of an NGO that contains a massive amount of agricultural data. There may be a discussion about whether to make it public and freely available. Maybe to farmers and to the public, but free of charge to a private company? I do not think so, because of the value and potential benefit of the data, and the cost that the community has borne to collect and to curate the data. The materiality of the digital, as we discussed earlier, is expensive.

Another example is free access to data about their ancestors for any individual in, for instance, the archives in the Netherlands. This is all classic and very popular, and it is a beautiful thing. Now, if you start using these data for more than just genealogical reasons, e.g. by combining them with the DNA database, or start selling products, it is a different story. For, who has the power to reorganise all those data in a sort of profitable way? Companies yes, but not single individuals. It would be naïve just to open everything and welcome anyone to take advantage of the data resources made accessible. It is very expensive to collect and to curate all those data. So some of the value should go back into the community. A private company should pay an extra fee to use data from public archives. That funding could go back to the archives, and more archival resources could be made available to the public. There is an argument, at least here in the UK, in favour of opening databases and archival material from the government for entrepreneurial use by start-ups, which could have the opportunity to find ways of monetising the data. This is fine, but the data will not be used only by start-up companies. This is why I think something like a *freemium* model would be much preferable: free for individuals, more expensive for companies, and the bigger the company the higher the fee may be.

EDITORS: This kind of regulation does not exist yet in the public sphere. It would be very difficult to implement.

FLORIDI: Perhaps, but it is not unprecedented. Companies that for instance have financial data and sell them, put online only some bits of data that is free for you to see. But if you want 'the real thing' then you have to pay. It is not a model that everybody knows, and it is not in use with public databases, but it provides a good example.

As to open data, remember that the open data movement started as a political movement in terms of transparency of government. However, it soon became something else, once it became coloured by financial, and not longer political, interpretations. Initially, the open data discussion was about making the government more transparent and hence more accountable: one may see where the money goes, who does what, and who is responsible for what kind of program, for example. From there, the goals slowly morphed into commercial (re-)use by start-ups for innovation, and things ended up with potential exploitation by big companies. What transparency is there in giving access to let's say records of hospitals to a private company? It is not about transparency. It is not about a start-up. It is about a company that is taking a huge advantage of costly public records. I am very much in favour of it, but I would add a price.

EDITORS: Could you elaborate on the term '*hyperhistory*'? It might be that the hyperhistorical result in more or other tasks and goals for the archival community.

FLORIDI: Hyperhistory is a neologism I introduced in a recent book called *The Fourth Revolution – How the Infosphere is Reshaping Human Reality*. It is based on a simple idea. Time has classically been divided into prehistory and history. Prehistory refers to any stage of human development where there exists no means of r ecording the present for future consumption; in particular, societies without writing. Prehistory ended around 6000 years ago in Europe and China where –

simultaneously – writing was invented. Since then, we have been increasingly living in information societies. Only a few people in the Amazonian environment still live prehistorically. Today, if we describe history as our interaction with information and communication technologies from writing to press, printing, the radio, cinema, mass media and so forth – we have exponentially increased our dependence on these technologies to the point where, with the advent of digital, our *dependence* on technology is *absolute*. In some corners of the world we live more '*historically*' than ever before. The wellbeing of the individual and the welfare of the society is no longer just historically *related* to ICTs it is *dependent* on them, *hyperhistorically*. In many places in the world and certainly in Europe the proper functioning of society depends on digital infrastructure. This means they could be subject to cyber-attack, a good test to understand whether you live *hyperhistorically*. History has become even more historical than ever before. And the entire world is heading the same way. There is no '*end of history*', because history is a technological concept not a political one.

Allow me now to speculate for a moment, as if we had all the means and possibilities of changing the world in one go. We then could provide in more anchoring and more stability in all the informational fluidity or liquidity of today. What we need is more anchoring in this liquidity. It is fine and great to have all this fluidity and, for example, all the fake news. It is fine that there is freedom of speech. However, we have opened a kind of Pandora's Box, in the sense that now there are two billion voices on Facebook that can say whatever they want. That is freedom of speech, and that is a good thing. But they can also pollute, and then we end up not knowing anything. If any of them drops one piece of plastic in the sea you can imagine what happens. If each of them drops one piece of fake news in the infosphere, we no longer know what to believe, the noise obliterates the signal.

Where do we establish a little bit of cleaning and re-anchoring? I do not believe in not allowing people to talk, but I do believe in contributing to the conversation with plenty of good information. So, you start cleaning by sort of out-spacing the negative elements. It might be science fictional and speculative, but imagine the following. The archival community should be openly and seriously vocal about all the silly things people are saying and communicate: we have the documents, here is the authentic version and, this is how it is or went. However, that is not in the understanding of, for instance, the library and information science community. They do not think in those terms. The risk is to think more in terms of being a warehouse. Using the hyperhistory terminology, in a world that is becoming increasingly dependent on the digital, the crucial question is: who is providing the 'balancing act', who is keeping the infosphere clean? I do not mean to dump all this on the archival community and ask it to save the world. There are a lot of professions there, like teachers, scientists, educators at all levels and others that should do a better job. They are all called to contribute to the problem management in the space of information. But at the moment, the archivist profession seems noticeable for its absence.

EDITORS: Apart from this not being active enough, we also might have a more structural, or more institutional, issue here. In the paper world there was mainly the government. Now there are a lot of companies and private organisations that,

structurally and institutionally, decide how things are and how things work. When you see these developments, governmental archives seem to become less relevant. The question would be: how do we cope with this?

FLORIDI: A problem that I face in another context looks at the same issue but from a quite different angle. It concerns the proprietary nature of databases. Let me give examples of two companies, Amazon and Apple. Neither of them would be immediately identified with education or health. But in terms of, for example, reading abilities, Amazon probably knows more about how the world reads than anyone else. Because Amazon has huge amounts of data about all the e-books. It knows where people stop reading, where people have to read the page twice and which words they actually click because they do not know the meaning of the word. This is a treasure that should be used, but it is proprietary. I do not know whether Amazon is considering exploiting these data. Apple, with the iWatch and the iPhone, is probably the biggest collector of health-related data in the world as we speak, and it keeps growing. It says it is not in the data mining business. Yet Apple owns these data and I am not sure it is going to share them with, for example, the World Health Organization.

This brings us to the point that, I think, addresses your question: are we to ask digital companies, and who should ask? Or are we to push them, and say for instance: you are having access to the free databases of the government, therefore we should have free access to the databases of your results? Where is the mutual exchange of value here? Now consider that many companies, like Facebook, Twitter, Microsoft or Google, do collaborate with universities for research purposes. In this, they may give access to some of their data. But even Twitter, which is quite famous for sharing its databases (and therefore their archive), does it in limited ways. This means that you are not always certain that the data you get are fully representative. It can get unclear and confusing as to how much you can do with the data reliably. Or speculate for a moment about an imaginary day when Facebook decides to close down. What would it do with its data? If it were to donate its data to a government, would you trust that government with those data more than you trust Facebook? I'm really not so sure.

We have a current project on data donation, supported by Microsoft. We are exploring the possibility, at the European level, to devise a simple code of practice to facilitate the donation of medical records by individuals after their death, a bit like organ donation. Personally, I would like to donate my medical data for research to the National Health Service. I will be dead by then, so privacy is not an issue for me. This example goes in the direction of your question, in terms of mutual interactions concerning who owns which data, and for what purpose the data are used. Private companies or organisations donating their archives for public use is, of course, not an entirely new phenomenon. What is new, is the staggering dimension. When you take the example I gave of Facebook, that is two billion people connected, and an enormous amount of records. It is staggering. It would be great to see all stakeholders taking steps towards a mutual interaction between what is public, governmental, propriety, and invite companies to contribute to the welfare of the world by sharing more of their data.

EDITORS: A book by Jaron Lanier: *Who owns the future* (2013) is about regulating the personal possession of data.

FLORIDI: I had several interactions with him, as we are members of the same committee on the European General Data Protection Regulation. I'm afraid I disagree with what I take it to be his vision. It is natural to think that data are a property and therefore that data usage can have a price that will be regulated by the market. But the truth is that personal information is not about what the 'market' should regulate. It is about social preferability of what we want to do. Data become useful and valuable only when hugely aggregated. That is why I keep stressing that what makes the difference here is the amount of data. Lately I checked the value of my profile, I think it was a service provided by the Financial Times. It was less than the value of a song on iTunes. I am sure that is the case for most people. Nobody cares about a grain of sand. Everybody cares about the beach. So I would argue exactly the opposite: precisely because money is left out we can care about personal information, because it is not a possession, but the priceless information that constitutes one's personal identity. If we start attaching one dollar of worth to a personal profile, we're done. If there is no price attached to those personal data, then the use of those data is not regulated by the forces of the market.

EDITORS: There is also another issue concerning ownership of information, at least in Dutch law. You cannot really own the information 'itself'. You only own the carrier: the floppy, the (piece of) paper, the disk drive or whatever. When you would regulate ownership of information 'itself', you could also regulate responsibilities concerning the information. What is your point of view on this?

FLORIDI: There is very little information that we own strictly speaking, unless you have the copyright on something. In that case you basically own the information, the content. Anything else is not covered by law, there is no contract.

I think it is important to understand that the private owes badly to the public. The private has taken huge advantage, rightly so and legally so, of public resources. It would be great to see the private put back into the 'common good' some of the value extracted from such resources. I am talking about data here, just data. Because when it comes to software services, it is the other way round, the public is taking an enormous advantage of the private. Take all the services we have that are free, just because someone somewhere is providing them in exchange of personal data and attention. The governments should have done that, but they did not. This brought us the current situation of data exploitation. I can image a world in which a company says to a government: you are using my software for free, so I am using your data for free.

Data donation, to me, is part of the solution for a better future. However, there has to be a shift in culture. It is a small shift from the philanthropic donation of money for the public good – which is not uncommon – to the donation of data. Today, data is the valuable resource. There must be a switch in perspective.

EDITORS: By way of finishing this interview, could you elaborate on what archivists and/or the archival community should or shouldn't do, maybe on the ethical side, given all the developments in archives and data we discussed?

FLORIDI: That is quite an open question, and it is a big one. I would like to address a more general ethical issue relating to current digital developments, that of self-determination. In this regard, the questions to ask are: how much do we want to have digital media empowering people to determine their lives and be well informed? Can people still decide to have a life apart from the constraining power of the profiling techniques? What can we do if the interest of the profiling or monitoring entity becomes mainly to influence and predict, so to manipulate the behaviour of the individual? We should not be too paranoid, they might just want to sell me another fridge, which is not the end of the world. But what if they stop selling me fridges and instead want to sell me ideas about what world I should live in? This is the scary bit, because what we have at the moment is basically a 'fingers crossed strategy'. At the moment, the immense power they have is exercised in a mostly benevolent way. But we are relying on hope, that nothing goes wrong, and that is not reassuring.

It is the self-determination and the autonomous individual that is at stake. Of course, basic trust is important in the present world. But should we just rely on trust? Should we not also have some constraints, accountability, liability, expectations? In English we use trust also as a verb. 'I trust you' means that I believe strongly that you are a good person, that you mean well, and will do your best to deliver on expectations. That is also the way we are trusting digital companies. I trust they will not do anything harmful. I trust that if they do not behave well is because it is a mistake, not a plan. Yet this looks to me like a weak strategy. What I do doubt is whether trust is a successfully strategy here; it is like leaving the door unlocked, and trust that no one comes in to steal anything. It is an unsafe kind of trust.

One of the final, defining questions concerning the next step of information society is: where should we support trust with further, social, legal, political frameworks? As we speak, the major actors are realising that they have to be good citizens. They are no longer playing the sort of 'we-are-neutral', 'we are not involved', 'we just give people what they want' kind of game. They are also given big fines by the European Union, and they might start listening. The question here is, how do we start changing and getting on with implementing the good side of citizenship. That means not only trust, but also playing according to the right rules and taking responsibility. In short, the biggest challenge in front of us is the governance of the digital.

I am an optimist, so be careful when I say that this is the way things will go. More pessimistically, perhaps I should say that *if* we have a good future for our information society, then that is the way forward – which does not mean that we will take it. It is more like: if we want a good information society, which is socially preferable and something that we would sign out for, it is a society in which big corporate actors in the information world play their role as good corporate citizens.

It is not just about political governance anymore. We are past that stage of human history. It went as far as it could: it is modernity. Thinking that the State is going to fix all our digital problems it is not being in touch with the twenty-first century. The governance of the digital will have to involve all stakeholders.

The massive liquidity has to be counterbalanced by some *anchoring*. This reminds me of the logo of Aldus Manutius, the most famous Italian publisher and one of the founders of the modern book era. He had a logo with the anchor and a dolphin and the motto '*Festina Lente*' which means 'Go fast, slowly'. This logo, I think, stands for a good balance for information society. The dolphin represents liquidity, the anchor represents the necessary grounding. Because it is not just about liquidity of the digital, it is also about grounding the governance of the digital. We need a counterbalance to the excessive liquidity of our society. I think it would be a great motto and icon for the society we want to develop.

Literature

Floridi, L. (2002). *Sextus Empiricus - The Recovery and Transmission of Pyrrhonism*. Oxford: Oxford University Press.

Floridi, L. (2014). *The Fourth Revolution - How the Infosphere is Reshaping Human Reality*. Oxford: Oxford University Press.

Glaudemans, Arnoud, Rienk Jonker and Frans Smit (2015). The right people and the right forces need to get together, Interview with Prof. Luciano Floridi, *Archievenblad, L-F, 01/2015*. Retrieved from http://www.youblisher. com/p/1048348-Archievenblad-uitgave-van-de-Koninklijke-Vereniging-van-Archivarissen-in-Nederland/

Lanier, Jaron (2013). *Who owns the future?* London: Allen Lane.

Yeo, G. (2008). Concepts of Record(2), Prototypes and boundary objects. *The American Archivist, 71*, 118-143.

BIO

Biographies

Geert-Jan van Bussel has studied medieval history, business administration, business informatics and archival studies. He is a certified archivist. He is assistant professor at HvA University of Applied Sciences Amsterdam. From 2012-2016, he was professor Digital Archiving and Compliance at that university. He is, with *Van Bussel Document Services,* an independent consultant, auditor and researcher. He is visiting lecturer at several universities in the Netherlands and Europe. He was president of the Special Commission for Archives, a commission of the Council of Culture, the most important advisory body on culture for the Dutch government. He is a (keynote) speaker on many seminars and conferences, mostly on the effects and influence of information processing and information management on people's work. He published extensively on digital archiving, accountability, and Enterprise Information Management. In 2001, he has been awarded the prestigious NMA Award, an award of excellence from the Dutch Association for Information Management to acknowledge his merits for the Document-, Workflow- and Record Management market.

Wolfgang Ernst is Full Professor for Media Theories in the Institute for Musicology and Media Studies at Humboldt University in Berlin since 2003. Having been academically trained as a historian (PhD) and classicist (Latin philology and Classical Archaeology) with an ongoing interest in cultural tempor(e)alities, he grew into the emergent technology-oriented "German school" of media studies and His academic focus has been on archival theory and museology, before attending to media materialities. His current research covers media archaeology as method, theory of technical storage, technologies of cultural transmission, micro-temporal media aesthetics and their chronopoetic potentials, and sound analysis ("sonicity") from a media-epistemological point of view.

Books in English: Digital Memory and the Archive (2013); Stirring in the Archives. Order from Disorder Stirring in the Archives (2015); Chronopoetics. The temporal being and operativity of technological media (2016); Sonic Time Machines. Explicit Sound, Sirenic Voices and Implicit Sonicity in Terms of Media Knowledge, Amsterdam (2016)

Luciano Floridi is Professor of Philosophy and Ethics of Information at the University of Oxford, where he is also the Director of the Digital Ethics Lab of the Oxford Internet Institute. Still in Oxford, he is Distinguished Research Fellow of the Uehiro Centre for Practical Ethics of the Faculty of Philosophy, and Research Associate and Fellow in Information Policy of the Department of Computer Science. Outside Oxford, he is Faculty Fellow of the Alan Turing Institute (the national institute for data science) and Chair of its Data Ethics Group; and Adjunct Professor ("Distinguished Scholar in Residence") of the Department of Economics, American University, Washington D.C.

His research concerns primarily Information and Computer Ethics (aka Digital Ethics), the Philosophy of Information, and the Philosophy of Technology. Other research interests include Epistemology, Philosophy of Logic, and the History and Philosophy of Scepticism. He has published over a 150 papers in these areas, in many anthologies and peer-reviewed journals. His works have been translated into many languages, including Arabic, Chinese, Dutch, French, German, Greek, Hungarian, Italian, Japanese, Lithuanian, Persian, Polish, Portuguese, Russian, and Spanish. His lifetime project is a tetralogy (not his term) on the foundation of the philosophy of information, called Principia Philosophiae Informationis.

Fiorella Foscarini is an associate professor in the Faculty of Information at the University of Toronto, Canada. In 2014-16, she taught in the Department of Media Studies at the University of Amsterdam, The Netherlands. She holds a PhD in Archival Science from the School of Library, Archival and Information Studies at the University of British Columbia in Vancouver. Before joining academia, she worked as senior archivist for the European Central Bank in Frankfurt am Main, Germany. Prior to that, she was Head of the Records Office and Intermediate Archives at the Province of Bologna, Italy. In her teaching and research, she uses diplomatics, rhetorical genre studies, and information culture concepts to explore issues related to the creation, management, and use of records in organizational contexts. She is co-editor in chief of the Records Management Journal.

Anne J. Gilliland is Professor and Director of the Archival Studies specialization in the Department of Information Studies, Director of the Center for Information as Evidence, Graduate School of Education & Information Studies, and a faculty affiliate of the Center for Digital Humanities at the University of California Los Angeles (UCLA). She is also the director of the Archival Education and Research Initiative (AERI), a global collaborative effort amongst academic institutions that seeks to promote state-of-the-art in scholarship in archival studies, broadly conceived, as well as to encourage curricular and pedagogical innovation in archival and recordkeeping education locally and worldwide.

She is a Fellow of the Society of American Archivists and recipient of numerous awards in archival and information studies. She is an Honorary Research Fellow of the Centre for Global Research, RMIT University in Melbourne.

Her research and teaching relate broadly to the history, nature, human impact and technologies associated with archives, recordkeeping and memory, particularly in translocal and international contexts. Her recent work has been addressing recordkeeping and archival systems and practices in support of human rights, recovery and daily life in post-conflict and diasporic settings; the role of community memory in promoting reconciliation in the wake of ethnic conflict; bureaucratic violence and the politics and nature of metadata; digital recordkeeping and archival informatics; and research methods and design in archival studies.

Arnoud Glaudemans works at Streekarchief Gooi en Vechtstreek in Hilversum as supervisor of the information management at the six affiliated govermental organisations. He studied philosophy and archival studies in Amsterdam. As a member of the archival advisory committee of the Dutch association of municipalities (VNG) he is actively involved in the development of various practical tools in information management (e.g., appraisal, quality management).

Juan Ilerbaig holds a MISt from the University of Toronto (2011) and a Ph.D. in the History of Science and Technology from the University of Minnesota (2002). For the past few years he has taught courses as a sessional instructor at the University of Toronto, in both the iSchool (Archives and Records Management Concentration) and the Institute for the History and Philosophy of Science and Technology (History of Evolutionary Biology). His research interests focus mostly on three areas: the interactions between record keeping and the practice of science, particularly in the life sciences; the application of a genre perspective in both archival science and the history of science; and the history and philosophy of archival concepts and theories. Current work in progress focuses on the genre systems used by Charles Darwin in his natural history research and on the uses of geological and pictorial metaphors in archival thinking and theory.

Charles Jeurgens is professor of archival studies at the University of Amsterdam (since 2016) and advisor at the Dutch National Archives (since 2009). He published extensively on issues of appraisal and selection, colonial and postcolonial archival cultures. He studied history and archivistics and did a PhD in the history of 19th century infrastructural planning in the Netherlands. He worked as editor of archival sources of the Batavian-French period at the Institute of Netherlands History in The Hague and he was municipal archivist of Schiedam (1994-1999) and Dordrecht (1999-2009). He was professor of archivistics at Leiden University between 2004 and 2016.

Rienk Jonker has been working as archivist since 1981. After ten years at the Centrale Archief Selectie Dienst in Winschoten, an agency of the Ministry of the Interior, he returned to the municipal archive of the city of Groningen, what later became part of the RHC Groninger Archieven, as archival inspector and later municipal archivist with the instruction to advise about and accompany the transition to the digital world from an archivist perspective. Since 2006 he has been working for the municipality of Leeuwarden with almost the same assignment. In 2008 he became the municipal archivist of Leeuwarden.

His main areas of interest are the basics of archival science, information architecture, metadata, information processing, appraising and appraisal of records, terminology, the digitization of the work environment and the management and preservation of digital records. From 1999 until 2011, he has on almost monthly basis provided colleagues with information about developments that touch the horizon of the archivist. From 2004, he maintains his own website on records and information management and archives under the motto there is nothing news under the sun (www.labyrinth.rienkjonker.nl). In 2009, he received the Van Wijnpenning from the Royal Association of Archivists in the Netherlands (KVAN) as a token for his work.

Eric Ketelaar is Professor Emeritus at the University of Amsterdam, where from 1997 to 2009 he was Professor of Archivistics in the Department of Mediastudies. As an honorary fellow of his former department he continues his research which is concerned mainly with the social and cultural contexts of records creation and use. From 1989-1997 he was General State Archivist (National Archivist) of The Netherlands. From 1992-2002 he held the chair (part-time) of archivistics in the Department of History of the University of Leiden. Eric Ketelaar was visiting professor at the University of Michigan (Ann Arbor), Gakushuin University (Tokyo), the University of Toronto and Monash University (Melbourne), where he continues to be involved as a Senior Research Fellow. From the foundation, in 2001, of Archival Science, he was one of the editors-in-chief. Since 2014 he is a member of the Editorial Board.

Giovanni Michetti is Assistant Professor of Archival Science at Sapienza University of Rome. His research area is focused on contemporary and digital archives. His main research interests are records management, description models and digital preservation. He has been involved in national and international projects on digital preservation, including ERPANET (Electronic Resource Preservation and Access Network) and CASPAR (Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval), both funded by the European Commission. He is currently leading researches within the InterPARES Trust project. He is heavily involved in standardization processes as the Chair of the Subcommittee "Archives and Records Management" and Vice-Chair of the Committee "Documentation and Information" in UNI, the Italian Standards Organization. He is also the Italian representative in a few ISO Working Groups on archives and records management.

Martijn van Otterlo obtained his PhD (artificial intelligence, A.I.) from the University of Twente (Netherlands, 2008) with a dissertation on expressive knowledge representation in machine learning from evaluative feedback. He published two books on such adaptive learning algorithms (2009 IOS Press; 2012 Springer, together with Dr. Wiering). Martijn has worked on robotics, vision and language and held positions in Freiburg (Germany), Leuven (Belgium) and Nijmegen (The Netherlands). His second research interest, which arose from his expertise in A.I., concerns the ethics and implications of adaptive algorithms on privacy, surveillance and society. He has served as committee member and reviewer for dozens of international journals and conferences on machine learning, data science and artificial intelligence. In his current position at the Vrije Universiteit Amsterdam (The Netherlands) he combines data science and ethics with his third interest: libraries. He currently studies the digitalization of physical, public libraries, and also the ethical consequences of datafication of library processes and services. More information can be found at http://martijnvanotterlo.nl

Frans Smit is Information Governance Officer at the Dutch Municipality of Almere, and Teacher of Archival Science at the University of Amsterdam. Educated as a historian, he has been working in fields like software engineering, archives, libraries and information policy departments as a developer, governance officer, manager and consultant. He publishes regularly in various journals, predominantly about cross-boundaries between information disciplines. He was co-editor of the S@P-Yearbook on archival inspection "Profiteer, profileer, prioriteer" (2013). He is a member of the archival advisory committee of the Dutch Association of Municipalities (VNG) as well as a member of the Records Management Expert Group (RMEG) of the International Council on Archives (ICA). He is a consultant and a trainer for among others the National Archives of Indonesia (ANRI, 2012-2013), the Dutch Archiefschool (2014-) and the Royal Association of Archivists in the Netherlands (KVAN, 2017-).

Jacco Verburgt studied philosophy and theology in Amsterdam, Leuven, Berlin, and Rome. He is a researcher at Tilburg University, the Netherlands. His current research focuses on Aristotle, Aquinas, Kant, and Hegel, but includes applied science topics too. He also is an editorial board member of Critique, which is an online publishing platform (see https://virtualcritique.wordpress.com). He teaches philosophy (especially courses on history of philosophy, philosophical anthropology, general and applied ethics, and philosophy of science) at various institutions of higher education in the Netherlands.

Geoffrey Yeo is an Honorary Senior Research Fellow in the Department of Information Studies at University College London (UCL), United Kingdom. Before joining UCL in 2000, he worked as an archivist for the Corporation of the City of London, for St Bartholomew's Hospital, London, and for the Royal College of Physicians of London. He has also worked as a freelance archivist and records manager, and as a consultant to the International Records Management Trust, participating in records management and educational projects in The Gambia, Ghana, Botswana and Uganda. In 2010 he was Visiting Professor at the School of Library, Archival and Information Studies at the University of British Columbia, Vancouver, Canada.

His research interests include conceptual understandings of records; perceptions of the origins and scope of record-making and record-keeping; records' contextualisation and description; and relations between records and the actions of individuals and organisations. He has published widely on description and on conceptual understandings of records and is a frequent speaker on these and related topics at international academic and professional conferences. His published work won the Society of American Archivists Fellows' Ernst Posner Award in 2009 and the Hugh A. Taylor Prize in 2013.

Atlantis

The ultimate platform for collection management and portal construction

Atlantis is a flexible and reliable system for collection management and portal construction.

If you wish to manage your heritage resources in a professional and responsible way and wish to offer your customers and users an optimal heritage experience, Atlantis is the perfect choice.

DEVENTit
Developers & Inventors in IT

www.deventit.nl

Experience the heritage

OUR EXPERTISE IS TO **DIGITALIZE YOUR COLLECTIONS!**

- Notarial archives
- Old-court archives
- Church archives

From collection to connection!
Karmac would like to help you to process the first step. We offer you the possibility to digitalize your old-court archive, notarial archive, slides, newspapers, population registers, microfiches etc. in a secure and trusted environment. This offers you a lot of benefits like: reducing the use of the originals; widening public accessibility or connection with other collections.

Curious to the benefits? Please contact us and we will make your collections digital. You can reach us by:

- +31 (0)320 - 28 69 59
- info@karmac.nl
- karmac.nl

karmac

Archives in Liquid Times aims to broaden and deepen the thinking about archives in today's digital environment. It is a book that tries to fuel the debate about archives in different fields of research. It shows that in these liquid times, archives need and deserve to be considered from different angles.

Archives in Liquid Times is a publication in which archival science is linked to philosophy (of information) and data science. Not only do the contributors try to open windows to new concepts and perspectives, but also to new uses of existing concepts concerning archives. The articles in this book contain philosophical reflections, speculative essays and presentations of new models and concepts alongside well-known topics in archival theory.

Among the contributors are scholars from different fields of research, like Anne Gilliland, Wolfgang Ernst, Geoffrey Yeo, Martijn van Otterlo, Charles Jeurgens and Geert-Jan van Bussel. This book includes interviews with Luciano Floridi and Eric Ketelaar, in which they reflect on key issues arising from the contributions. The editors are Frans Smit, Arnoud Glaudemans and Rienk Jonker.