Archives in Liquid Times

Edited by: Frans Smit, Arnoud Glaudemans and Rienk Jonker

design - www.absoluutdesigners.com

printed by - GTV Drukwerk Project Management bv

ISBN EAN 978-90-71251-45-0
Stichting Archiefpublicaties, 's-Gravenhage 2017

This publication has been made possible by:
Archiefschool/Hogeschool van Amsterdam (cover 2, page 6)
DEVENTit B.V. (page 6, 328)
Doxis Informatiemanagers (page 1, 6)
Picturae (page 2, 6)
Karmac Informatie & Innovatie B.V. (cover 3, page 6)
De Ree archiefsystemen (page 6)

Archives in Liquid Times

edited by

Frans Smit
Arnoud Glaudemans
Rienk Jonker

Jaarboek 17
Stichting Archiefpublicaties
s@p

Table of Contents

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.

> *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]

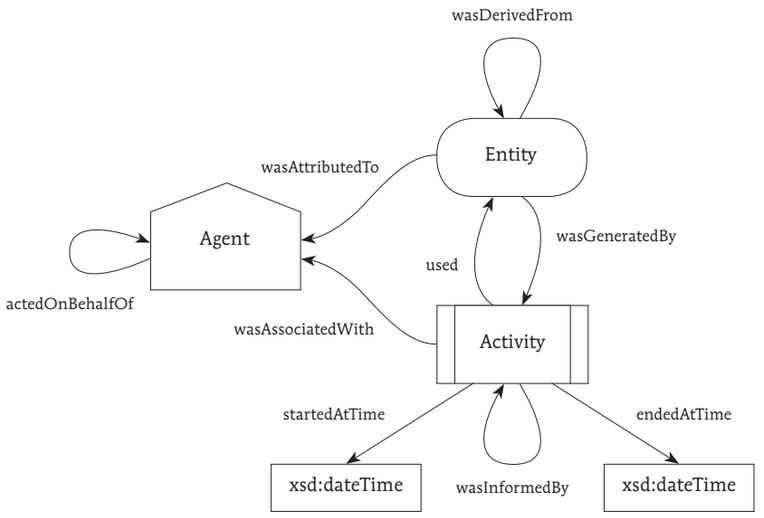

Figure 1. The basic classes and properties of the PROV ontology.

[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

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.

ISBN/EAN 978-90-71251-45-0
Jaarboek 17, Stichting Archiefpublicaties