ISSN 0204–2061. KNYGOTYRA. 2014. 63

RECONCEPTUALIZING RECORDS, THE ARCHIVE AND ARCHIVAL ROLES AND REQUIREMENTS IN A NETWORKED SOCIETY

Anne J. Gilliland | Department of Information Studies
University of California Los Angeles
GSE&IS Building, Box 951520
Los Angeles, CA 90095-1520
E-mail: Gilliland@gseis.ucla.edu

Records created through institutional and personal activity are the primary concern of archives, regardless of the form or media of those records. This paper focuses on the implications of the networked society for how such records are created, as well as how they are defined and appraised by archivists, and the kinds of searching and uses to which they might be subjected. It argues that in responding to these implications, archives will have to move beyond the custodial and institutional mindset that has dominated the field for centuries and instead embrace a network orientation. It proposes and provides examples of several potential conceptualizations of "networked records": the multi-provenance bureaucratic record and the record created by the crowd; the metadata record; the extra-institutional record, the transinstitutional and the transjurisdictional record, and the mobile or itinerant record; the stitched-together record; the mined, mapped and compiled record; and the implied or inferred record.

The paper also identifies several dissemination and access challenges faced by archivists when networked records and their metadata accumulate into a virtual "archival corpus". These challenges include the ways in which that corpus might be mined, forensically analyzed, cross-compiled, found lacking, augmented and otherwise searched or mapped to the benefit or detriment of organizational, scholarly, community and personal interests. After identifying a set of human rather than asset, data or task-centred principles that it is argued should inform archival activities such as appraisal, description and dissemination, the paper concludes that traditional appraisal techniques alone are unable to cope with these challenges. Instead the field working together with researchers in information retrieval (IR) should focus on innovating in the area of archival information storage and retrieval ("archival

IR") in ways that can exploit the networked creation and uses of records and other forms of primary data and their metadata, and respond to and protect against potential vulnerabilities, particularly those relating to privacy and security, that might be exposed by such developments.

KEYWORDS: *archives, information retrieval, networked records.*

INTRODUCTION

Dynamic information and record-creating technologies, shifting trends in historical and cultural scholarship, the burgeoning community archives movement, social media archiving, the push for open data, and human rights and social justice concerns have all presented compelling challenges for archives to re-conceptualize key archival principles, concepts and behaviors over the past few decades.[1] However, as important a factor as each of these concerns is in its own right, intertwined with and propelling each in various ways is the phenomenon that arguably presents the greatest impetus of all for archival reconceptualization and technological development – and that is the networked society.

Records created through institutional and personal activity are the primary concern of archives, regardless of the form or media of those records. This paper focuses on the implications of the networked society for how such records are created, as well as how they are defined and appraised by archivists, and the kinds of searching and uses to which they might be subjected. It argues that in responding to these implications, archives will have to move beyond the custodial and institutional mindset that has dominated the field for centuries and instead embrace a network orientation. It proposes and provides examples of several potential conceptualizations of "networked records": *the multi-provenance bureaucratic record* and *the record created by the crowd; the metadata record; the extra-institutional record*, *the transinstitutional* and the *transjurisdictional record,* and *the mobile* or *itinerant record; the stitched-together record; the mined*, *mapped* and *compiled record;* and *the implied* or *inferred record*.

The paper also identifies several dissemination and access challenges faced by archivists when networked records and their metadata accumulate into a virtual "archival corpus." These challenges include the ways in which that corpus might be mined, forensically analyzed, cross-compiled, found lacking, augmented and otherwise searched or mapped to the benefit or detriment of organizational, scholarly, community and personal interests. After identifying a set of human rather than asset, data or task-centred principles that it is argued should inform archival activities such as appraisal, description and dissemination, the paper concludes that traditional appraisal techniques alone are unable to cope with these challenges.

Instead, the field working together with researchers in information retrieval (IR) should shift its focus and resources to innovating in the area of archival information storage and retrieval ("archival IR") in ways that can exploit the networked creation and uses of records and other forms of primary data and their metadata, and respond to and protect against potential vulnerabilities, particularly those relating to privacy and security, that might be exposed by such developments.

BACKGROUND

Networking *per se* is not anything new to the archival field, either administratively or technologically. Indeed, the notions of administrative, procedural and documentary contexts that are applied in modern diplomatics to explain the bureaucratic structures, workflows and various kinds of relationships between documents (e.g., second copies, different versions, series of documents)[2] can all be depicted as forms of networks with nodes or nexuses of action, agents and relationships,[3] and have been discernable for as long as formal bureaucracies have existed. Archivists intuitively understand these kinds of networks and their impact not only on the production but also on the control and compilation of the record, especially in complex and geographically distant situations such as colonial administrations or collaborative science.[4]

When computer networking began to be implemented for military, research and business purposes, particularly from the 1970s onwards, as surely as it was

1 GILLILAND, Anne J. *Conceptualizing twenty-first-century Archives.* Society of American Archivists, 2014.

2 DURANTI, Luciana. *Diplomatics:* New Uses for an Old Science. Society of American Archivists, Association of Canadian Archivists, and Scarecrow, 1998.

3 These are also the basis for the entity-relationship models that are applied in the Australian Recordkeeping Metadata Schema (RKMS). MCKEMMISH, Sue; ACLAND, Glenda, REED, Barbara, and WARD, Nigel. Describing Records in Context in the Continuum: The Australian Recordkeeping Metadata Schema. *Archivaria*, Fall 1999, vol. 48, p. 3–37.

4 See, for example RAMAN, Bhavani. *Document Raj*: Writing and Scribes in Early South India. Chicago, 2012; WAREHAM, Evelyn. From Explorers to Evangelists: Archivists, Recordkeeping and Remembering in the Pacific Islands. *Archival Science*, 2002, vol. 2, p. 187–207; WARNOW-BLEWETT, Joan; GENUTH, Joel; and WEART, Spencer R. IAIP Study of Multi-institutional Collaborations: Phase III: Ground-based Astronomy, Materials Science, Heavy-Ion and Nuclear Physics, Medical Physics, and Computer-mediated Collaborations. Report No. 1: Summary of Project Activities and Findings. Project Recommendations. American Institute of Physics, 1999 [accessed 10 September 2014]. Access through Internet: <http://www.aip.org/history/pubs/collabs/phase3rep1.htm>; SANDS, Ashley, BORGMAN, Christine L., WYNHOLDS, Laura and TRAWEEK, Sharon. Follow the Data: How Astronomers Use and Reuse Data. In *Proceedings of ASIST 2012*. Baltimore, MD, 2012 [accessed 10 September 2014]. Access through Internet: <https://www.asis.org/asist2012/proceedings/Submissions/341.pdf>.

taken up by other areas of information production, dissemination and use, it was also taken up by archives. Archival descriptive systems have been networked since the 1980s and 1990s, first through bibliographic utilities such as the Research Libraries Information Network and then the Web.[5] Over the past two decades, the exchange and mapping of archival descriptive metadata has been facilitated by the development of national and international descriptive standards and more recently through linked data and the Semantic Web.[6] Consortia such as ICARUS and Monasterium, and portals such as Europeana and Matricula[7] have been created to collate, promote and exhibit content from different repositories and are vehicles for sharing resources and expertise and developing mutually beneficial strategies and standards. Many other archives have at least experimented with some kind of collaborative strategy for creating and/or collecting documentation of particular phenomena, or for developing joint virtual exhibitions.

Nevertheless, none of this is truly responding to the concept and pervasive effects of the networked society on the record, on the archive, and on how that archive might be used. A networked society is one where social, economic, political and cultural life is facilitated by and indeed is created through ubiquitous connectivity via digital information and communication networks. It is a reality in many nations in the world, with the entire globe moving rapidly in this direction. Much of the canon of archival ideas and their implementation in practice remain firmly based in non-network thinking, however, running the risk of the professional management of archives being relegated in the popular mind to content generated in the pre-digital networking world. The roots of this can be traced back to some very fundamental ideas in archival science that derive from the traditional and still primary role of archives as institutional record-keepers. These ideas continue to reflect historical standalone bureaucratic structures and notions of centralized authority and control and do not translate well to a more diverse and widespread use and user base, nor do they protect equally the interests of all who are invested in or subjects of those records. Among these ideas are the following:

- that records are works of an official or legal nature that may be used as evidence or proof. A broader definition would be that they are the byproducts of organizational and personal activity. They have three components: content, context and structure.[8]
- that records are generally unpublished and unique;[9]
- that those records should be physically and intellectually retained and collectively described together according to the entity that generated them, i.e., their official provenance;
- that such *fonds* are created and accumulated *by* and not *about* that entity; and,

- that that entity is usually conceptualized as a single organization, authority, office, function or individual, albeit one that may take on different forms or names as its historical role unfolds. That is to say, it has a single provenance.

Little of this reflects the reality of bureaucratic records creation today, and certainly not of the broader cultural record as it is being digitally created, transmitted, linked and curated both inside and outside the auspices of the mainstream archival profession. Archivists engage in a certain professional schizophrenia in that they continue to promote these principles at the same time as they are increasingly aware of their conceptual and practical limitations and constraining aspects in the digital realm and in particular in a networked society. Longstanding tensions between the public records and research orientations underpinning institutional archives and collecting repositories have been a source of debate and compromise in developing description and access standards that can handle bureaucratic, scholarly, cultural and community conceptions of what might be considered and treated as a record. However, there have also been eloquent and substantial challenges on several other fronts that have considerable resonance when we contemplate how to conceptualize a record as well as any kind of virtual and networked archive or corpus of records in a networked society. In 1981, F. Gerald Ham initiated a professional dialog on the "post-custodial era" when he argued that the abundance of computer-generated records and copies thereof was forcing archivists out of their introspective proclivities and excessive proprietaryness towards their holdings and into a more proactive, less custodial role.[10] His argument also challenged the idea of the uniqueness of records, underscoring how easy it was to produce digital copies. This challenge to uniqueness was taken up by others also. James O'Toole listed several additional ways or contexts in which uniqueness might be contemplated in respect to archives and that are highly relevant to how we contemplate records and the iformation they contain in a networked context today. These reach beyond undercomplexified assertions about "the uniqueness of records" that are, as Ham pointed out, challenged in

5 GILLILAND, Conceptualizing Twenty-first-century Archives, ibid.

6 GILLILAND, ibid.

7 ICARUS, <http://icar-us.eu/>; Monasterium, <www.monasterium.net>; Europeana, <http://www.europeana.eu/>; Matricula, http://matricula-online.eu/ [accessed 10 September 2014].

8 PEARCE-MOSES, Richard. *A Glossary of Archival Terminology*. Society of American Archivists, 2005 [accessed 10 September 2014]. Access through Internet: <http://www2.archivists.org/glossary>.

9 PEARCE-MOSES, ibid.

10 HAM, F. Gerald. Archival Strategies for the Post-custodial Era. *The American Archivist*, Summer 1981, vol. 44, no. 3, p. 207; ACLAND, Glenda. Managing the Record Rather than the Relic. *Archives and Manuscripts*, 1992, vol. 20, no.1, p. 58–59; BEARMAN, David A. Record-keeping Systems. *Archivaria*, 1993, vol. 36, p. 16–37.

a world of abundant and easily generated digital copies to include: "the uniqueness of information in information in records; the uniqueness of the processes which produce records; and the uniqueness of the aggregations of documents into files."[11] The InterPARES 2 Project also acknowledged the problem of uniqueness as a defining characteristic of a record in a networked environment, given the possibility of the existence of multiple simultaneously-generated identical digital original documents. Like O'Toole, it extended prior theory (in this case, that of contemporary archival diplomatics) with regard to the many and inter-dependent contexts of those documents:

> *Context* shifts the analysis away from the record itself to the broader structural, procedural, and documentary framework in which the record is created and managed. The identified elements of context … include the record's ***juridical-administrative context***, its ***provenancial context***, its ***procedural context***, its ***documentary context***, and its ***technological context***.[12]

In similar fashion, the conceptualization of a single or authoritative creator and provenance has been challenged through the proposition of co-creatorship, and multiple simultaneous and parallel provenance:

> These propositions argue that traditional notions of provenance are oversimplified. With their emphasis on a single creating entity, [they] fail to acknowledge that multiple parties with different types of relationships to each other can be involved in the genesis of records. The propositions maintain, for example, that subjects as well as creators of records should be acknowledged as participants in that genesis and that archivists have an ethical imperative to pursue descriptive mechanisms for representing both creator and co-creator worldviews and experiences, and supporting diverse users' needs and concerns, within and relating to a given community of records.[13]

For many years, archivists have been focused on the management of "born-digital" (a.k.a. electronic) records and have developed models, technologies and protocols for their capture, ingest, preservation and description (although notably there has been considerably less emphasis on searching and retrieval of the archived born-digital record). More recently, they have been seeking to apply or extend these approaches to records storage and management in the Cloud. Expanded definitions for a record that called out some aspects that usually were tacitly understood or were readily apparent in the physical world were found to be necessary to identify and characterize the record in the digital world and were formulated accordingly. For example, InterPARES 1 defined an "electronic record" as being:

... like its traditional counterpart, ... a complex of elements and their relationships. It possesses a number of identifiable characteristics, including a fixed documentary form, a stable content, an archival bond with other records either inside or outside the system, and an identifiable context. It participates in or supports action, either procedurally or as part of the decision-making process (meaning its creation may be mandatory or discretionary), and at least three persons (author, writer, and addressee) are involved in its creation ... these same or similar elements are present explicitly or implicitly in electronic records."[14]

Gilliland-Swetland and Eppard defined electronic records as:

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. Identifying the boundaries of such intellectually complex objects and then moving those objects forward through time and through migrations without compromising their authentic status is a significant issue.

Records are temporally contingent -- they take on different values and are subject to different uses at different points in time. Records are also time-bound in the sense that they are created for a specific purpose in relation to a specific time-bound action.[15]

But again, such definitions remained limited in application to certain categories of records explicitly understood as 'electronic records' and have rarely been systematically applied to the entire universe of records (i.e., not just the bureaucratic, but also the broader human and cultural record), whether digital or physical, tangible or intangible.[16] Moreover, we also know that the current paradigm

11 O'TOOLE, James M. On the Idea of Uniqueness. *The American Archivist*, Fall 1994, vol. 57, p. 632–658.

12 DURANTI, Luciana, ed. The InterPARES Project: The Long-term Preservation of Authentic Electronic Records: The Findings of the InterPARES Project. *ArchiLab*, 2005, p. 27.

13 GILLILAND, A. J. Conceptualizing Twenty-first-century Archives, p. 29. See also HURLEY, Chris. Parallel Provenance: (1) What, If Anything, Is Archival Description? *Archives and Manuscripts*, 2005, vol. 33, no.1, p. 11–45; HURLEY, Chris. Parallel Provenance: (2) When Something Is Not Related to Everything Else. *Archives and Manuscripts*, 2005, vol. 33, no. 2, p. 52–91; KETELAAR, Eric. Sharing: Collected Memories in Communities of Records. *Archives and Manuscripts*, 2005, vol. 33, no. 1, p. 50; GILLILAND, Anne J. Contemplating Co-creator Rights in Archival Description. *Knowledge Organization*, 2012, vol. 39, no. 2, p. 340–346.

14 DURANTI, L. The InterPARES Project, p. 25.

15 GILLILAND-SWETLAND, Anne J. and EPPARD, Philip B. Preserving the Authenticity of Contingent Digital Objects: The InterPARES Project. *DLib Magazine* (July/August 2000) [accessed 10 September 2014]. Access through Internet: <http://www.dlib.org/dlib/july00/eppard/07eppard.html>.

16 The notable exception being the Australian

that is used to manage the creation and use of active records and to appraise or assess those that should be retained by archives over the long-term fails to grapple successfully with "born-networked" records and with associated questions such as what is the archival record and what its metadata, when is the archival record, where is the archival record, and whose is the archival record?[17]

NETWORKED NOTIONS OF RECORDS

Indeed, we could identify many new post-physical ways to conceive of how, when and where the record is created or otherwise manifested in a networked environment and reorient archival thinking and practices accordingly. For example, we could conceptualize the record in the following ways:

The multi-provenance bureaucratic record and the record created by the crowd:

As already discussed, several archival scholars in recent years have pointed out that the dominant archival conceptualization of provenance fails to acknowledge the complex of parties that are often responsible for or participate in the creation of a record. They point out that this perpetuates existing bureaucratic power structures and elites and renders others who participate in the production of the record as mere subjects rather than co-creators with rights in those records. Despite such acknowledgment of the over-simplification of provenance and its human consequences, the descriptive standards community has resisted building more complexity into descriptive standards regarding provenance, and thus IR also has a strong dependency on this under-complexified notion.[18] Born-networked, multi-provenancial records, such as those generated by organizational or scientific research collaborations, or within large-scale social media or other Web 2.0, or even the smarter but potentially less equitably organized and accessible coming Web 3.0 environments, make it impossible to continue to ignore this issue, no matter how problematic it might be for archival arrangement and description practices.

The metadata record:

Electronic records management research has demonstrated the ways in which digital records are a composite of both content and their contextual metadata that has accumulated over the life of that record.[19] This might have seemed to be a somewhat academic conceptualization until recently. However the scandals that erupted as a result of disclosures about the U.S. National Security Agency and other governments intercepting mobile phone communications have surfaced some very interesting questions regarding mobile telephony about what comprises a re-

cord. A record in the context of a digital communication has often been equated to the content of a text message or phone call and may be legally protected. What comprises the metadata for that text message or phone call (which is often equated to everything about a person's phone calls or messages that is not the content, e.g., patterns of calling, duration of calls, routing over cell networks, and destination of calls or messages), may not be recognized as part of the same record and be similarly legally protected. Indeed, such metadata, which is often distributed across many different parts and places of a communications network, is currently being exploited by intelligence and security agencies that argue that it is not a record in its own right or even a protected part of the content that is recognized as a record.

In terms of how and where that record is created and managed, and under whose auspices, we could come up with some other clusters of conceptualizations of the record:

The record that does not exist simply within a single institution's custody or jurisdiction:

a. The extra-institutional record:

Examples of such records would include those stored in a commercial Cloud, onshore or offshore, often on the same servers as records of other organizations and individuals and potentially subject to multiple jurisdictional and security claims. If we think again about the record being created by mobile telephony, although senders and recipients view the record on their own devices, they are not readily empowered to archive it, and archives even less so. Moreover, copies of the record or its metadata that have been transmitted over networks are often being captured and being stored by a telecommunications provider who also might be subject to

conceptualization of records within the Records Continuum, which is extended into RKMS. MCKEMMISH et al., Describing Records in Context in the Continuum, ibid.

17 ACKER, Amelia. When is a Record? A Research Framework for Locating Electronic Records in Infrastructure. In *Research in the Archival Multiverse*. Monash: anticipated publication, 2015; GILLILAND, Anne J. Archival Appraisal: Practising on Shifting Sands. In *Archives and Recordkeeping*: Theory Into Practice. Facet Publishing, 2013, p. 31–61.

18 GILLILAND, A. J. Contemplating Co-creator Rights in Archival Description, ibid.

19 GILLILAND, A. J., et al. Investigating the Roles and Requirements, Manifestations and Management of Metadata in the Creation of Reliable and Preservation of Authentic Electronic Entities Created by Dynamic, Interactive and Experiential Systems: Report on the Work and Findings of the InterPARES 2 Description Cross Domain Group, Part VI, in International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records. Associazione Nazionale Archivistica Italiana [accessed 10 September 2014]. Access through Internet: <http://www.interpares.org/ip2/book.cfm>.

various legal requirements about how long those copies must be kept and whether they must be opened upon demand to a government's security agencies.

b. The transinstitutional or the *transjurisdictional record*:

Examples of these might be the kinds of records that are generated through various types of collaboration within and across public and private sector interests. Government, corporate and scholarly activities are among those that regularly create records in this way (and they may use Cloud services for storing and accessing these records), often to find that the records that are created become orphaned at the end of a collaboration because no archival authority is responsible or designated for archiving such networked transinstitutional activities. Records saved by individual collaborators might selectively end up in the institutional archive for a given collaborator, and thus be distributed in an ad hoc way across multiple institutional repositories. Moreover, such records may be subject to competing ownership and access claims because of the different legal jurisdictions and private and public institutions involved with their creation.

c. The mobile or itinerant record:

This is a record that isn't tethered to where and when an individual or group is working – a person can create it everywhere s/he goes. S/he carries it with him or her, and s/he may have certain discretion over whether or not it gets archived. Examples might include the contents of a business or personal mobile phone (texts, calling records, photographs and video, social media interactions) as well as those of solid state drives and flash memory.

Alternative conceptualizations in terms of how a record might be discerned, evaluated and used when its existence is not immediately apparent

Here is where we see some of the real potential, as well as threats, of a networked approach. The following are three related examples:

a. The "stitched- together" record:

This could also perhaps be called the distributed or the diasporic record, or even the record of virtual traces. Very different parties are interested in these phenomena – for example, scholars who trace the routes of diasporic and migratory populations and individuals are interested in the traces of those individuals that might show up in historical immigration records, passenger lists, boarding house registers, aid agency reports, and newspaper business advertisements that have been digitized and/or digitally described by various archives and put online.[20] Another example that is very similar and yet very dissimilar at the same time, might be a national security agency that is trying to track the movements of individuals suspected of being involved in terrorist activity – whether that be through active mobile phone

records and metadata, or through such traces as recent immigration records, passenger lists, and hotel registrations.[21] Yet another example might be to support the documentary needs of victims of human rights abuse, refugees, other emigrants, and migrant workers who may pass through many points of documentation around a region or around the world as they move and who have need of those records to establish rights, citizenship, residence, or eligibility for workers' or veterans' benefits or healthcare.[22] Such examples illustrate why so many people are interested in linking together documentary traces that can provide a bigger picture than a single archive might afford. There are many ways in which this stitching-together could be done beyond present efforts, by exploiting a record's metadata across its life and by using various pattern matching and inferencing techniques, but these have a variety of implications not only for scholarship and genealogy, but also for personal privacy and national security, and these implications must be researched in concert with new techniques for information retrieval from and across corpora of records.

b. The mined or mapped or compiled record:

This could also be thought of as a latent record that can be actualized by mining a corpus of records or mapping or compiling across one large, or multiple corpora in order to draw a picture of an individual or an event, or to detect patterns that would be impossible to discern from individual records or corpora. One of the most prominent examples of this is the Twitter Archive. In 2010 the Library of Congress in Washington, D.C. entered into a controversial partnership with Twitter and social data provider Gnip to build and preserve an archive of tweets. In 2013, the Library of Congress justified its decision to preserve or "archive" the Twitter Archive as follows:

> As society turns to social media as a primary method of communication and creative expression, social media is supplementing and in some cases supplanting letters, journals, serial publications and other sources routinely collected by research libraries.
>
> Archiving and preserving outlets such as Twitter will enable future researchers access to a fuller picture of today's cultural norms, dialogue, trends and events to inform scholarship, the legislative process, new works of authorship, education and other purposes.[23]

20 For example, BALD, Vivek. *Bengali Harlem and the Lost Histories of South Asian America*. Harvard, 2012.

21 For example, the European Data Retention Initiative. European Commission, Home Affairs, Data Retention [accessed 10 September 2014] <http://ec.europa.eu/dgs/home-affairs/what-we-do/policies/police-cooperation/data-retention/index_en.htm>.

22 For example, personal medical records: <http://www.raconline.org/rural-monitor/mivia-program-electronic-health-records/>, [accessed 10 September 2014].

23 LIBRARY OF CONGRESS. Update on the Twitter Archive at the Library of Congress,

In other words, the Library was arguing that social media can capture a more complete archive of certain facets of society than was ever previously possible-indeed, filling in the blanks of what archives historically have either not chosen to acquire at all, or have selectively weeded out. Both the numbers of tweets and the overall volume of the archive are enormous, currently growing at over half a billion a day. The Twitter feed is acquired in real-time (i.e., without any lag time between tweeting and ingestion) – and although this may be necessary in order to ensure that tweets are captured, it raises new questions about privacy and the value of hindsight that archival laws such as decades-long delays in materials being transferred to archives in certain ways addressed. It is also being acquired without going through any appraisal process (i.e., no selection mechanism is used, for example, to identify only tweets associated in some way with America or to impose any kind of archival value judgment on the continuing value of the tweets).

The Library of Congress argues that, "It is clear that technology to allow for scholarship access to large data sets is lagging behind technology for creating and distributing such data. Even the private sector has not yet implemented cost-effective commercial solutions because of the complexity and resource requirements of such a task."[24] To encourage such development of information retrieval capabilities, and to address the impossibility of manually describing the contents of the archive, data mining rather than archival description of the chronologically-organized Twitter Archive is being conducted by Gnip. While this data-centric approach offers the possibilities of being able to do many new things with the contents of the archive, something is also likely being lost. Manually-created archival description traditionally would bring a value-added component to the dissemination and retrieval process, that would include incorporating a broader contextualization of content and also taking measures to protect sensitive content or the interests of individuals mentioned in the content who might in some way be vulnerable.

c. The implied or inferred record:

Another kind of latent record that is somewhat more difficult conceptually to grasp is the record that is present through its absence. It could be argued that this is a record of the personal or social imaginary—it is the record a scholar, genealogist, plaintiff or survivor wishes were there but just isn't, but depending upon the robustness of the other material in the corpora, one might infer that it originally was there, or it should have been created but wasn't. An example of this approach has been attempted by an historian, Matthew Connelly, working with a team of computer scientists. Connelly asked whether big data mining techniques could be applied to the contemporary holdings of the U.S. National Archives which, he said

"had more holes than a donut factory" as a result of the classification and consequent 30-year closure of many government documents. By learning everything he could about how records are created, maintained and released to the public and also compiling and analyzing traces, patterns and anomalies in digitized copies of available paper records he and his team found that it was possible to identify records that were missing because they had been withheld or even to discern the outlines or draft of a document that was not present. Connelly has also begun to analyze metadata records in order to discern patterns of racial profiling that might have been used by government agencies.[25]

CONCLUSION: THE NEED FOR A HUMAN RATHER THAN ASSET, DATA OR TASK-CENTERED ARCHIVAL IR

Given such scenarios archivists need to accept that they are increasingly unable, and indeed, will not be able to afford to, continue to exercise control over the circumstances of creation or production of the record. Traditional records management activities such as records retention scheduling and appraisal (the act of selecting what will be archived and what disposed of) are struggling in the network society, and indeed will likely become obsolete. Even if it were possible for archivists to discern and to obtain a mandate to appraise the various manifestations of digital records, digital forensics (which includes computer forensics, network forensics, forensic data analysis, and mobile device) have clearly demonstrated that it is wellnigh impossible to eliminate, beyond recovery, all traces of a networked digital record and its linkages. It is also increasingly difficult to discern what of a granular series of communications or other digital traces archivists could – intellectually, physically, and at a realistic economic cost, selectively eliminate. Moreover, archivists in individual physical repositories may find that they will never be in a position to take physical custody or even legal ownership of the networked record for which they might be responsible or expected to curate as part of a virtual archive.[26] Such curation will instead likely involve virtual capture and preservation, and then ensuring the capacity to retrieve, construct, or reconstruct, what will be

January 4, 2013 [accessed 10 September 2014]. Access through Internet: <http://blogs.loc.gov/loc/2013/01/update-on-the-twitter-archive-at-the-library-of-congress/>.

24 LIBRARY OF CONGRESS, ibid.

25 CRAIG, David J. The Ghost. *Columbia Magazine*, Winter 2013-14, p. 17–23.

26 For example, even though the Library of Congress has invested vast amounts of money in acquiring and maintaining the Twitter Archive, Twitter remains the owner of the content.

a profoundly networked and vastly more extensive and hopefully more inclusive bureaucratic, human, cultural and community record.

So what should archivists be doing? They need to re-orient themselves to a world not only made up of network-born and network-discernable records, but also of the resulting accumulations of networked archival corpora, whether those be the by-products of scientific endeavor or of social media. Instead of spending their energies and limited financial and technological resources on trying to reduce the bulk of the records through appraisal, they should be focusing on ways in which different kinds of records can productively and effectively be extracted from these corpora. IR has made very few substantive inroads into the archival world. Nevertheless, it is clear from the Twitter example that if archivists are not prepared to move into this realm, with or without assistance from colleague in computer and information science, commercial developers will do so instead. Moreover, the Twitter example clearly illustrates why automated retrieval directly from an extensive corpus of born-networked materials may be the only way to find material, and some of the new ways in which such a corpus might be mined and contents "stitched-together."

As the preceding discussion suggested, archives are full of unknowns – their extensiveness and latent granularity mean that there is always new material and new knowledge to be uncovered. Moreover, it is the unknown, or the not previously viewed, that is often precisely what the historian, journalist, lawyer, human rights activist or genealogist wishes to uncover. This is one of the areas where archival applications of IR might also contribute a new perspective back to the parent field of IR – which historically has been focused on achieving a match between what is sought by a user and what is known to exist and has been described according to particular rules in the corpus. In the archival case, however, the user may be interested in using IR techniques to establish what does not exist in the corpus in order to make inferences about why it is not there. It is rather like how astronomers build their knowledge of the universe – in part relying upon the dense maps of what is currently known of the universe as a way to throw into relief the dark spaces where what they contain is unknown. By examining how known objects are possibly being affected by objects as yet unknown or invisible, and by otherwise hypothesizing about why there is dark space, that dark space itself increasingly becomes populated by what must, or might be there.

At the same time, however, archivists should renew their focus on some of the other functions traditionally performed by appraisal such as ensuring that information relating to personal privacy or corporate or national security is not accidentally exposed, that vulnerable individuals discussed in the records are protect-

ed, and that records are retained that were useful and usable. Most IR is centred around the data or the asset being retrieved, or the task the end-user is trying to undertake. However, such models are insufficient for the scale and sensitivity of human concerns that emerge when entire corpora of digital networked records are retained and cross-searched, compiled or mined. Elsewhere I have proposed a platform that foregrounds several "ethical" acts-- *Acknowledging, Respecting, Enfranchising, Liberating and Protecting*--that do not appear in the mainstream rhetoric of information retrieval or indeed of information organization theory and practice but that I argue should lie at the centre of archival activities such as appraisal, description and access, particularly in such a networked and granularly documented world.[27] Among the recommended principles would be that:

- Archives will acknowledge both the creators and the co-creators/subjects of records when appraising, describing and making accessible those materials.
- To the fullest extent possible, archives will consult with the creators and co-creators/subjects of archival materials when appraising, arranging materials, developing descriptions and making decisions about access and disclosure.
- Archives will strive to identify and implement mechanisms for enhancing the visibility, findability and usability of archival material relating to communities and experiences that have historically been under- or inequitably represented or rendered invisible through archival descriptive practices.
- Archives will acknowledge and respect the belief systems and traditional cultural expressions of the creators and co-creators/subjects of archival materials when developing archival descriptions and online access systems.
- Archives will work to ensure that their appraisal and descriptive practices or access and disclosure processed do not expose or exploit those who are vulnerable to suppression, appropriation, violence, discrimination or other oppressive or traumatising acts, or re-traumatise them. This includes future generations that might be vulnerable on the basis of what is contained in the archives.[28]

If we could leave the corpus intact and lossless in terms of contextual relationships and metadata (i.e., ingesting and/or maintaining it without any reduction through appraisal), but operationalize these principles through the clever exploitation of metadata created throughout the life of the record and records creation and keeping processes, as well as the design of archival storage and retrieval systems,

27 GILLILAND, Anne J. *Acknowledging, Respecting, Enfranchising, Liberating and Protecting:* A Platform for Radical Archival Description. Paper. Radical Archives Conference, New York University, April 2014.

28 GILLILAND, A. J.; and MCKEMMISH, Sue. The Role of Participatory Archives in Furthering Human Rights, Reconciliation and Recovery. *Atlanti*: Review for Modern Archival Theory and Practice, in press, vol. 24.

we could retain a record that is richer in social detail than anything previously retained by humanity, but one hopes, build in safeguards against inappropriate exploitation. At the same time, the application of these principles in archival IR, together with employing IR to help users to find previously unknown and possibly "smoking gun"–type documents; establishing the meaningful absence (as opposed to the presence) of documents; and exploiting multiple types and sources of metadata—might well find wider application in other domains applying IR techniques such as litigation support systems, news retrieval, audiovisual archives, data mining, and digital asset management.

Literature

1. ACKER, Amelia. When is a Record? A Research Framework for Locating Electronic Records in Infrastructure. In *Research in the Archival Multiverse*. Monash, anticipated publication 2015.
2. ACLAND, Glenda. Managing the Record Rather than the Relic. *Archives and Manuscripts*, 1992, vol. 20, no. 1, p. 57–63.
3. BALD, Vivek. *Bengali Harlem and the Lost Histories of South Asian America*. Harvard, 2012. 320 p. ISBN 978-0674066663.
4. BEARMAN, David A. Record-keeping Systems. *Archivaria*, 1993, vol. 36, p. 16–37.
5. CRAIG, David J. The Ghost. *Columbia Magazine*, Winter 2013-14, p. 17–23.
6. DURANTI, Luciana. *Diplomatics:* New Uses for an Old Science. Society of American Archivists, Association of Canadian Archivists, and Scarecrow, 1998. 198 p. ISBN 978-0810835283.
7. DURANTI, Luciana, ed. The InterPARES Project: The Long-term Preservation of Authentic Electronic Records: The Findings of the InterPARES Project. *ArchiLab*, 2005.
8. GILLILAND, Anne J. *Acknowledging, Respecting, Enfranchising, Liberating and Protecting:* A Platform for Radical Archival Description. Paper. Radical Archives Conference, New York University, April 2014.
9. GILLILAND, Anne J. Archival Appraisal: Practising on Shifting Sands. In *Archives and Record-keeping: Theory Into Practice*. Facet Publishing, 2013. 288 p. ISBN 978-1856048255.
10. GILLILAND, Anne J. *Conceptualizing twenty-first-century Archives*. Society of American Archivists, 2014. 336 p. ISBN 1-931666-68-7. 336 p.
11. GILLILAND, Anne J. Contemplating Co-creator Rights in Archival Description. *Knowledge Organization*, 2012, vol. 39, no. 2, p. 340–346.
12. GILLILAND, Anne J., et al. Investigating the Roles and Requirements, Manifestations and Management of Metadata in the Creation of Reliable and Preservation of Authentic Electronic Entities Created by Dynamic, Interactive and Experiential Systems: Report on the Work and Findings of the InterPARES 2 Description Cross Domain Group, Part VI, in International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records. Associazione Nazionale Archivistica Italiana. Access through Internet: <http://www.interpares.org/ip2/book.cfm>.
13. GILLILAND, Anne J.; and MCKEMMISH, Sue. The Role of Participatory Archives in Furthering Human Rights, Reconciliation and Recovery. *Atlanti*: Review for Modern Archival Theory and Practice, in press, vol. 24.
14. GILLILAND-SWETLAND, Anne J.; and EPPARD, Philip B. Preserving the Authenticity of Contingent Digital Objects: The InterPARES Project. *DLib Magazine* (July/August 2000). Access through Internet: <http://www.dlib.org/dlib/july00/eppard/07eppard.html>.
15. HAM, F. Gerald. Archival Strategies for the Post-

custodial Era. *The American Archivist*, Summer 1981, vol. 44, no. 3, p. 207–316.
16. HURLEY, Chris. Parallel Provenance: (1) What, If Anything, Is Archival Description? *Archives and Manuscripts*, 2005, vol. 33, no. 1, p. 110–145.
17. HURLEY, Chris. Parallel Provenance: (2) When Something Is Not Related to Everything Else. *Archives and Manuscripts*, 2005, vol. 33, no. 2, p. 52–91.
18. KETELAAR, Eric. Sharing: Collected Memories in Communities of Records. *Archives and Manuscripts*, 2005, vol. 33, no.1, p. 44–61.
19. LIBRARY OF CONGRESS. Update on the Twitter Archive at the Library of Congress, January 4, 2013. Access through Internet: <http://blogs.loc.gov/loc/2013/01/update-on-the-twitter-archive-at-the-library-of-congress/>.
20. MCKEMMISH, Sue; ACLAND, Glenda; REED, Barbara; and WARD, Nigel. Describing Records in Context in the Continuum: The Australian Recordkeeping Metadata Schema. *Archivaria*, Fall 1999, vol. 48, p. 3–37.
21. O'TOOLE, James M. On the Idea of Uniqueness. *The American Archivist*, Fall 1994, vol. 57, p. 632–658.
22. PEARCE-MOSES, Richard. A Glossary of Archival Terminology. Society of American Archivists, 2005. Access through Internet: <http://www2.archivists.org/glossary>.
23. RAMAN, Bhavani. *Document Raj*: Writing and Scribes in Early South India. Chicago, 2012. 296 p. ISBN 978-0226703275.
24. SANDS, Ashley, BORGMAN, Christine L., WYNHOLDS, Laura and TRAWEEK, Sharon. Follow the Data: How Astronomers Use and Reuse Data. In *Proceedings of ASIST 2012*, Baltimore, MD, 2012. Access through Internet: <https://www.asis.org/asist2012/proceedings/Submissions/341.pdf>.
25. WAREHAM, Evelyn. From Explorers to Evangelists: Archivists, Recordkeeping and Remembering in the Pacific Islands. *Archival Science*, 2002, vol. 2, p. 187–207.
26. WARNOW-BLEWETT, Joan; GENUTH, Joel; and WEART, Spencer R. IAIP Study of Multi-institutional Collaborations: Phase III: Ground-based Astronomy, Materials Science, Heavy-Ion and Nuclear Physics, Medical Physics, and Computer-mediated Collaborations. Report No. 1: Summary of Project Activities and Findings. Project Recommendations. American Institute of Physics, 1999. Access through Internet: <http://www.aip.org/history/pubs/collabs/phase3rep1.htm>.

IŠ NAUJO PERMĄSTANT DOKUMENTUS, ARCHYVUS, JŲ VAIDMENIS IR REIKALAVIMUS TINKLAVEIKOS VISUOMENĖJE

Anne J. Gilliland

Santrauka

Institucinės ir asmeninės veiklos metu sukurti dokumentai turi būti archyvų veiklos dėmesio centre, nepriklausomai nuo šių dokumentų formos ar pateikimo būdo. Straipsnyje nagrinėjami tinklaveikos visuomenei būdingi pokyčiai, nulemiantys tokių dokumentų kūrimo formą ir būdus; analizuojama ir tai, kaip patys archyvarai pateikia, apibrėžia ir vertina šiuos dokumentus; tiriami galimi jų paieškos bei vartojimo būdai. Autorė teigia, kad, reaguojant į tinklaveikos visuomenėje vykstančią plėtrą ir jos implikacijas, archyvų veikloje turi įvykti tam tikrų pokyčių – nuo saugojimu ir instituciniu principu pagrįsto požiūrio, vyravusio šioje srityje ištisus amžius, būtina pereiti prie tinklaveikos pagrindu suformuotos pozicijos. Straipsnyje pateikiami potencialių tinklaveikos

dokumentų (angl. *networked records*) koncepcijų pavyzdžiai: daugialypės kilmės įstaigos dokumentas (angl. *the multi-provenance bureaucratic record*) ir masių dokumentas (angl. *the record created by the crowd*); metaduomenų dokumentas (angl. *the metadata record*); neinstitucinis dokumentas (angl. *the extra-institutional record*), tarpinstitucinis (angl. *the transinstitutional*) ir transjurisdikcinis dokumentas (angl. *transjurisdictional record*) bei mobilusis (angl. *the mobile*) arba keliaujantis dokumentas (angl. *itinerant record*); kartu sudarytas dokumentas (angl. *the stitched-together record*); apibrėžtas (angl. *the mined*), pažymėtas (angl .*mapped*) ir sudarytas dokumentas (angl. *compiled record*); numanomas (angl. *the implied*) arba išvestinis dokumentas (angl. *inferred record*).

Straipsnyje nagrinėjamos ir kai kurios sklaidos bei prieigos poblemos, su kuriomis susiduria archyvarai, kai besikaupiantys tinklaveikos dokumentai ir jų metaduomenys pavirsta virtualiu archyvų rinkiniu. Aktualu rasti būdų ir metodų visiems šiems archyvų tekstyno klodams tirti. Be to, būtina atlikti išsamią analizę bei kryžminį sudarymą, nustatyti trūkstamas vietas, papildyti, kitaip vykdyti paiešką ar pažymėti. Tokia veikla gali būti pravarti organizaciniams, moksliniams, bendruomenės ar asmenų interesams arba, priešingai, nesuteikti jokios naudos. Nustačius, kad svarbiau sutelkti dėmesį į visumą žmogiškųjų veiksnių, kurie, kaip manoma, turėtų padėti plėtoti tokią archyvų veiklą kaip vertės ekspertizė, aprašymas ir sklaida, užuot ir toliau koncentravusis į dokumentų rinkinių pridedamąją vertę, duomenis ar užduotis orientuotais principais, daroma išvada, kad vien tradicinių vertinimo metodikų nebepakanka šiems iššūkiams atremti. Būtina plėtoti bendradarbiavimą su informacijos paieškos specialistais (angl. *information retrieval*). Ne mažiau svarbu ieškoti archyvinės informacijos saugojimo ir paieškos naujovių nustatant būdus, kaip geriau panaudoti tinklaveikos dokumentų fondą; svarbu rasti dokumentų ir kitų pirminių duomenų bei jų metaduomenų panaudojimo optimalių būdų ir neatsilikti šioje srityje siekiant apsaugoti dokumentus nuo potencialių grėsmių, ypač susijusių su privatumu ir saugumu.

Įteikta 2014 m. rugpjūčio mėn.