RESEARCH ARTICLE

Open Social Knowledge Creation and Library and Archival Metadata

Dean Seeman and Heather Dean

University of Victoria, CA

Corresponding author: Dean Seeman (dseeman@uvic.ca)

Standardization both reflects and facilitates the collaborative and networked approach to metadata creation within the fields of librarianship and archival studies. These standards—such as Resource Description and Access and Rules for Archival Description—and the theoretical frameworks they embody enable professionals to work more effectively together. Yet such guidelines also determine who is qualified to undertake the work of cataloging and processing in libraries and archives. Both fields are empathetic to facilitating user-generated metadata and have taken steps towards collaborating with their research communities (as illustrated, for example, by social tagging and folksonomies) but these initial experiments cannot yet be regarded as widely adopted and radically open and social. This paper explores the recent histories of descriptive work in libraries and archives and the challenges involved in departing from deeply established models of metadata creation.

Keywords: Archives; Libraries; Metadata; Open Social Knowledge

Introduction

The fields of librarianship and archival studies have a long history of deeply collaborative and networked approaches to creating and circulating knowledge, from authority files for people, places, and subjects shared across institutions, to the first union lists, which compiled metadata across regional and national boundaries in order to provide unified access to information about cultural resources. But to what extent do these metadata ecosystems allow for open and social contribution? The distinct histories of archival and library metadata practice will be explored and differences in their approaches to metadata will be highlighted in order to help bring this issue into focus.

Archival Metadata

Within the past 30 years the archival community has established a collaborative and networked approach to creating and circulating metadata. In response to the advent of computers and the internet, the archival community developed content standards, XML schemas, and software for managing and describing archives (such as Access-to-Memory and ArchivesSpace) in the online environment, which are open source and community-developed.[1] Archival metadata, such as finding aids, are made available online through consortia regionally, nationally, and internationally, and the profession is increasingly employing linked open data as a means of facilitating discovery (such as the Social Networks and Archival Context Project (SNAC)).

Like librarians, archivists write metadata in accordance with professional standards. Perhaps, at first glance, the maxim of computer scientist Andrew Tanenbaum (1981) applies fittingly to archival description: 'The nice thing about standards is that you have so many to choose from [...]' (168). International standards for archival description were not in fact developed until the late twentieth century, and regional variations reflecting

---

[1] Throughout this article *archives*, rather than archive, is used to refer to 'materials created or received by a person, family, or organization, public or private, in the conduct of their affairs and preserved because of the enduring value contained in the information they contain or as evidence of the functions and responsibilities of their creator, especially those materials maintained using the principles of provenance, original order, and collective control; permanent records' (Pearce-Moses 2005).

different historical professional approaches can still be found. The International Council on Archives published the *General International Standard Archival Description* (ISAD(G)) in 1993 and the *International Standard Archival Authority Record for Corporate Bodies, Persons and Families* (ISAAR-CPF) in 1996. While many national standards, such as the American standard *Describing Archives: A Content Standard*, align with international standards, it is worth noting that regional variations can and do persist. The Canadian archival community's *Rules for Archival Description* (RAD), which is loosely based on the library community's *Anglo-American Cataloging Rules*, in fact pre-dates ISAD(G) by three years. While there are differences between ISAD(G) and RAD, the core six required fields are consistent.

These standards for archival description prescribe both schema and content; that is, the order and content of descriptive elements. XML schemas, such as *Encoded Archival Description* (EAD; 1998) and *Encoded Archival Context-Corporate Bodies, Persons, and Families* (EAC-CPF; 2011), are used in conjunction with descriptive and authority standards in order to render archival description machine-readable and facilitate the management and discoverability of archival metadata in a networked environment. Librarianship developed descriptive standards prior to the archival community largely because the benefits were immediately tangible: the same item held in multiple repositories could be described once and shared by many.

This pragmatic motivation for standardization does not have a corollary in archives, where each institution holds one-of-a-kind materials. Notably, the development of computers and the potential of making finding aids available online, and the requirement of consistency for interoperability, provided further impetus for the archival community to create standards. While archivists developed these standards with the intent of collaboration, it is worth noting that, given the relatively recent development and implementation of such standardization, the archival community is still aligning legacy metadata to such standards and making archival description broadly available online. For finding aids to truly be open, archivists will need to migrate existing metadata into the new platforms available. The challenge for the archival community is grappling not only with backlogs of undescribed archives, but also with the complexities of migrating legacy metadata.

In *Taking Our Pulse: The OCLC Research Survey of Special Collections and Archives*, Jackie Dooley and Katherine Luce (2010) found that half of archival collections have no online presence. One of the action items identified in the report is to make existing archival descriptions, or legacy finding aids, accessible online. Standardization and collections management software specifically designed for archives create the possibility of overcoming technological barriers to providing online access to archival description, although other challenges, such as migrating existing metadata, remain.

In addition to developing standards for making finding aids discoverable online, archivists and archival repositories have engaged with social media, social tagging and transcription, and crowd sourcing in order to further surface archival documents, as seen in Flickr Commons, blogs, YouTube channels, Instagram, and Twitter, as just a few examples. Archivists have experimented and advanced ways of incorporating new, more open and social functionality in finding aids, such as collaborative filtering and social navigation mechanisms and the ability for patrons to annotate finding aids (Hyry and Light 2002; Yakel et al. 2007). User studies have explored how online finding aids could be improved through clearer labelling of descriptive fields and better interfaces and navigation (Duff and Stoyanova 1998; Walton 2017). There is increasingly a call to address prejudiced language and perspectives in finding aids and the need to shift and share authority with communities of creators rather than situating descriptive authority purely within the profession (Drake 2016). These initiatives reflect a push within archives for collaboration not just within the profession, but between archivists and the communities in which they live and serve. In reimagining archival description, and its future, it is worth returning to the core purposes of finding aids and the principles underlying the work to describe archives.

An Aid for Finding (and So Much More)

Understanding archival metadata—such as that found in a finding aid—first entails an understanding of archives. Unlike published material, which exists in duplicate, archives are unique. There may be commonalities found across archives, such as authors (or creators), media (correspondence, diaries, photographs, and so on), and the activities and functions reflected in these documents, but in essence no two documents are alike. Unlike published items, such as books and journals, archives are generally not a discrete, bound object, but are comprised of aggregations of documents. Given that archives contain more than a single object, the task of archival processing (analogous to library cataloging) entails not just describing archives but also arranging them. These two qualities of archives—their singularity and multiplicity—have profound implications for archival metadata.

Finding aids, as the name suggests, are tools for locating archival material both physically and intellectually. The Society of American Archivists' *Glossary of Archival and Records Terminology* defines a finding aid as: 'A tool that facilitates discovery of information within a collection of records' (Pearce-Moses 2005). One could be tempted, then, to compare a finding aid to a bibliographic record for discovering a resource in a library, or even a table of contents or an index for navigating and locating information within a publication. A finding aid is none of these things despite the fact that it helps researchers to discover and locate information. Distilling the definition and purpose of a finding aid to this one literal meaning—an aid for finding—would overlook the broader importance such descriptions have, not only for researchers, but to the records and the creators they reflect. Indeed, in addition to facilitating access to archives, finding aids also serve to communicate the structure and context of archives, and in doing so, archival description plays a role in presuming the authenticity of records in an archives (Duff and Haworth 1997; MacNeil 2005). Unlike bibliographic description, which has access as its primary purpose, archival description has these two additional functions: 'to promote understanding of archival material by documenting its content, context and structure' and 'to establish grounds for presuming the authenticity of archival material by documenting its chain of custody, arrangement, and circumstances of creation and use' (Duff and Haworth 1997, 204).

An archival document is comprised of not only its content, but also its structure and context, and archival description captures and communicates these qualities. The content is the information expressed in a document, be it text or image, be it written on paper, parchment or another medium. The structure of a record 'relates to the physical and intellectual characteristics that define how a document was created and maintained' (Millar 2004, 7–8). The context of a record is the framework in which a record is created, used, and maintained. In their work, an archivist positions an archives within at least six contexts, including juridical-administrative, administrative, provenancial, procedural, documentary, and technological contexts (InterPARES 2005). The finding aid is essential for communicating this bigger picture—the structure and context—of a document.

The power of archives, the reason they inspire and compel research, is their ability to serve as authentic and reliable documentary evidence, to provide proof of decisions, opinions, recollections, or ideas (Millar 2004). However, for archives to effectively serve as evidence, archivists must arrange and describe fonds with consideration for the principle of *respect des fonds*—that is, the external integrity, or provenance, and internal integrity, or original order, of an archives. Provenance recognizes the organization or individual who created, accumulated, and/or maintained and used records in the conduct of their activities (InterPARES 2005); original order is the organization and sequence of records established by the creator of the records (Millar 2004). *Respect des fonds*, provenance, and original order constitute the core theoretical concepts underpinning the arrangement and description of archives; the archival community established these principles long before the development of the descriptive standards in use today.

The responsibility of archivists to identify and document the structure and context of an archives, as articulated in a finding aid, is scholarly work. Finding aids facilitate scholarship but that is not their only function. Often, processing archives entails identification and attribution of documents that, when done by a researcher, is often recognized as scholarly. Many archival repositories regard finding aids as publications and archivists as the authors of such publications. For large and/or complicated archives, arrangement and description can take months or years to complete. To do their work well, archivists have training in archival theory as well as in disciplines specific to understanding primary sources, such as diplomatics, paleography, and, increasingly, digital forensics. Consider the European professional context: in Italy, for example, the work of archivists to understand the context and the interrelationships of records—integral to ensuring the continuing authenticity of records—qualifies archivists as scientific researchers in legislation (Luciana Duranti, pers. comm., November 11, 2017). The process of arranging and describing a fonds is scholarly work, as is the product of this work. Increasingly, archivists are calling for recognition of the labour that such work entails, as well as for greater transparency into the work archivists do (Tansey 2015).

Authoring Archival Metadata

Archival description—like bibliographic description—seeks to be authoritative and objective. Even in deciding what to keep and what to discard, and how to describe that which is kept, the archivist is purportedly an impartial custodian, as articulated, for example, in the writings of British archivist Sir Hilary Jenkinson. Appropriately, such a position has come under criticism, with archivists questioning the capacity for neutrality and objectivity within the profession and calling for an awareness of biases (Caswell 2017; Cook 2001; Duff and Harris 2002; Drake 2016). One could also criticize the neutrality of a profession that

is overwhelmingly white, middle aged, and middle class, as found by the Society of American Archivists' A*CENSUS (2005). Perhaps a more open and social approach to writing archival description would allow for a multiplicity of perspectives and requirements, as modelled in projects such as the Mukurtu Content Management System, which facilitates access to digitized materials relating to indigenous communities using traditional knowledge labels and a sensitivity to cultural protocols.

A more open, collaborative, and social creation of archival description, that goes beyond current experimentation (such as social tagging and transcription) does not necessitate discarding the existing theoretical and intellectual framework in which archivists arrange and describe archives. The arrangement and description of an archives schematically based on research needs is problematic in that it has the potential to obscure the context in which a document is produced and to destroy its evidential value. An example from the University of Victoria (a case dating from a period prior to the professionalization of archivists in Canada and the establishment of descriptive standards) provides insight into the impact of competing research needs on processing an archives. In this example, the writings of an author were arranged—and continually rearranged—in response to research priorities. First the writings were organized into published and unpublished groupings. This enabled researchers to easily identify unpublished works and to subsequently publish these writings. Once this project was completed, researchers pointed out that the published and unpublished groupings were no longer relevant because everything had now been published. The fonds was rearranged two more times based on researcher requests: at one point the writings were put in chronological order to facilitate a biographer, and then at a later point the writings were arranged alphabetically by title of the works. What this example demonstrates is that research needs vary over time and from researcher to researcher. By placing research needs above those of the integrity of the archives, the employee dismantled the structure and context, the evidential value, of the fonds and, ultimately, an aspect of the research value of this particular writer's archives. While archivists cannot claim to be unbiased, they are professionally trained to be oriented first and foremost to the creator and the archives. A collaborative model bringing together the expertise of archivists in conversation with researchers and creators would prove a more efficacious alternative.

Archival description has changed over time, and undoubtedly arrangement and description practices and resulting finding aids will evolve in order to effectively capture and communicate the content, structure, and context of archives while meeting user needs. The University of Chicago's Mapping the Stacks project, which sought to identify and organize unprocessed archival collections that chronicle Black Chicago between the 1930s and 1970s, provides a useful instance of collaboration between archivists and scholars. This project is innovative as an example of cooperation and community-building between archivists and researchers. The challenge that Mapping the Stacks identified and resolved, at least on a small scale, is that of labour in the archives, of the incredible work it takes to provide access to large, modern archives. The project oriented subject experts (faculty and students) to archives and archival description and enabled participants to create finding aids for previously unprocessed and inaccessible material. The increasing complexity of managing contemporary archives will continue to impact how archivists approach the arrangement and description of such material, and the practicalities relating to the time-intensive labour of creating metadata will persist. Perhaps the most well-known response within the profession to the challenge of twentieth and twenty-first century archives is MPLP, or 'more product, less process,' first articulated in an article by Mark Greene and Dennis Meissner (2005), in which the authors advocate the use of minimal processing in order to reduce backlogs of unprocessed archives. Through minimal processing, archivists make a judgement call about when to provide less detail about a particular fonds in order to describe more fonds within a repository (in other words: a little bit of information about a lot of fonds is superior to a lot of information about a few fonds). This approach has the potential to create space that allows users to leverage their deep work with an archives in order to enhance archival descriptions. The challenge is whether or not users will indeed be motivated to contribute to archival descriptions. While there are certainly complexities to user-generated metadata, the intersection of existing professional expertise with an open and social environment for creating metadata provides one path forward in creating useful tools for discovering and interpreting archives.

Library Metadata
Like archives, libraries require metadata to describe their collections. Unlike archival description, context is not usually given primacy. This is partly due to the fact that libraries overwhelmingly collect published material (instead of documents), catalogue each individual item (as opposed to aggregations), and do so

at a scale at which noting individual context is impractical and usually not seen as necessary or valuable. When a library collects an item, it is generally identical to that which is collected by other libraries. Although markings, annotations, and publishing anomalies may set off an item as unique and noted as such (usually within rare book and special collections cataloguing), the entire library metadata model is driven by the idea that many libraries hold what amounts to the same thing.

Taking this into account, libraries have a long history of standardization and collaboration. Early approaches to systematic description were offered by Panizzi and Cutter in the nineteenth century (Denton 2007), but libraries soon saw the need for further collaboration, not only so that library users could expect a standardized approach to finding material across libraries, but so that libraries could share descriptions of resources to save time and money and reduce duplicated effort. As early as 1908, the American Library Association and the Library Association of the United Kingdom met to publish a common set of cataloguing rules (Denton 2007), and in 1961, lasting progress on collaboration was achieved when the International Conference on Cataloging Principles (ICCP) took place (Buizza 2004). National libraries met to establish international guidelines for bibliographic description; the 'Paris Principles,' as they are called, still form the basis of international cooperation today. The Americans and British worked together to create the Anglo-American Cataloguing Rules, based in part on the Paris Principles, and published it in 1967. The second revision of the Anglo American Cataloguing Rules (AACR2) followed in 1978 and incorporated ISBD, the International Standard Bibliographic Description, which was seen as a way to facilitate international communication for bibliographic information and convert it to machine-readable form (Taylor 2004). Besides international agreement on standards, the availability of new technologies saw the possibility of sharing records more efficiently. The Machine-Readable Cataloguing (MARC) format was created in the late 1960s by Henriette Avram for the Library of Congress (Gartner 2016). This encoding allowed for machine transmission and sharing of descriptions (as opposed to purchasing catalogue cards) and further entrenched the ideals of cooperation and standardization in the library world.

As the shared approach to creating bibliographic data became more sophisticated, a second type of data was considered: authority data. Authority records attempt to solve the problems of collocation (bringing together everything by an author or topic, no matter how they are represented on an item) and disambiguation (differentiating between two entities with the same name) by taking people, places, corporate bodies, subjects, series titles, and conferences and assigning one canonical, or 'authorized,' form of name for these entities. The authorized form is linked to alternative forms in an 'authority record' and this is also where these entities are described more fully (including profession, discipline, birth and death dates, etc.). This authority is used in particular fields in the bibliographic record to provide an authorized version of entities (in addition to transcribing how the entities are represented on the item itself). NACO (Name Authority Cooperative Project) was launched in 1977 to allow multiple libraries to contribute authority records. In order to transmit and encode authority data, *Authorities: a MARC Standard* was created in 1981 and its current iteration, MARC 21 Format for Authority Data is actively updated and maintained (Library of Congress 2018).

The landscape of networked collaboration today is a direct result of international standardization in decades past. Most libraries continue to employ MARC; some still use AACR2, while others have moved onto its successor, RDA (Resource Description and Access)—another international cataloguing content standard. Practically speaking, the Online Computer Library Centre (OCLC) Worldcat offers a database of over 380 million bibliographic records and has over 16,694 member libraries (OCLC 2016). Bibliographic records are added to the central OCLC database, and member libraries can upload, download and enhance existing descriptions. In North America, authority records are produced mostly as part of the NACO program, with member libraries contributing authority records about people, corporate bodies, and titles. This operates under the umbrella of the Program for Cooperative Cataloguing (PCC). Meanwhile, initiatives like Virtual International Authority File (VIAF) have brought multiple national authority files together to create an international authority file where no particular language or script is favoured and entities are assigned unique global identifiers.

Even in the 'non-traditional' world of digital object metadata description, the library has favoured international standards such as Dublin Core, IEEE's Learning Object Metadata (LOM), the Data Documentation Initiative (DDI), and the Library of Congress' Metadata Object Description Schema (MODS). The past century of library practice has demonstrated a continued trend to share, collaborate, and standardize on an international scale.

Does Collaborative Mean Open?

It is clear that the library world has a long tradition of collaboration, but does that mean it is also open? Thousands of cataloguers use international standards to create MARC records. They contribute bibliographic and authority records and in turn benefit from records created elsewhere. In the library environment, this network of contribution, collaboration, and re-use is firmly established.

However, as the broad array of standards above demonstrates, the expertise threshold for contributing to the traditional cataloguing environment is fairly high. Usually an apprenticeship of months to years is undertaken to make sure a cataloguer knows all the possible standards and is creating descriptions 'to standard.' Many of the standards are free to access (thanks to organizations like the Library of Congress and Library and Archives Canada), but other standards are not (such as the descriptive cataloguing standard RDA). Furthermore, contributing to collaboration in the library metadata environment usually requires access to a shared platform or service, which is usually mediated by a library instead of individual membership.

Although the path to collaboration is well set out, and librarians and library staff frequently collaborate with each other on metadata, traditional library metadata creation remains mostly closed due to technology, standards, and institutional barriers. But in what ways, if any, can the wider public become further involved in contributing to resource description?

The answer, in part, is that of course anyone can say anything about anyone or anything and make it public using an online platform. It is only when the library considers how to intersect these statements with 'official' library metadata creation and consumption that the question veers back into the library domain. As mentioned in association with archival metadata, Web 2.0 technologies have allowed for comments, tagging, and folksonomies in library catalogues and digital library platforms. Many other digital projects crowdsource transcriptions and other metadata related to a digital object. These initiatives allow open contributions to the 'official' metadata produced by libraries, but what are the implications?

Who Speaks for the Resource?

The history of library description and the library's use of international standards have created a situation where the barrier to metadata creation is high and the data is, on the whole, very standardized. However, this approach clearly limits what can be said about a resource and who gets to say it. Traditionally, the official library description of a resource carries with it the weight of authority, neutrality, and objectivity.

However, one person (or a few in the case of collaborative cataloguing) cannot claim to fully, completely, and neutrally describe a resource. According to Isabelle Boydens, even in the case of a single fact and of a single observer, an unequivocal informational representation of 'observable reality' is illusory (Bade 2011), and notions of objectivity disconnect libraries from their users. Socially constructed metadata, on the other hand, better reflects user terminology, enhances findability, improves serendipity, identifies the zeitgeist, and makes use of emerging vocabularies (Alemu et al. 2017). The library's metadata fails to represent the diverse worldview of its users. Its terms and controlled vocabularies are inadequate: they are outdated, missing context, and require that metadata specialists take on the task of 'mind reading' what they think is important (Shirky quoted in Alemu et al. 2012, 320). Furthermore, librarians and library staff are experts in the craft of cataloguing and classification, not necessarily domain experts equipped to semantically describe the content of material (Alemu et al. 2012). All of these critiques point to the need to facilitate broader contribution to metadata.

Practicalities

Taking into consideration these critiques, librarians may be convinced that more open metadata is necessary, but they are still constrained by practicalities. Systems need consistent data to work well, data needs to interoperate with other data, people describing things around the world require instruction on how to do so, users favour some kind of logical presentation of a library's resources, and libraries need to share descriptions to save on costs.

These legitimate requirements for internationalization and sharing (partly economic, partly to set consistent expectations for library users and systems) have a practical purpose, but perpetuate the idea that library metadata is authoritative. Libraries may realize that multiple interpretations are necessary, but are constrained by the high volume of material, purchased or acquired in both analogue and digital forms, that require description. Paralysis is not an option when people need to find resources. Even a thorough effort to be inclusive in metadata translates to less material catalogued (and leaves some material completely unfindable or hidden).

This marks a division between library description and scholarship in general: although library descriptions provide context and information about resources and entities, their purpose is, in the words of FRBR (Functional Requirements for Bibliographic Records), to find, identify, select, and obtain (IFLA 1997). The library first and foremost fosters discovery and facilitates scholarship. This is not to underplay the amount of research carried out to create a descriptive surrogate, but the research is not carried out for its own sake. In library description the goal is to practically organize a collection and enable discovery.

If the library sees metadata's purpose as primarily practical in nature, it would critique social and open contributions from the public on practical grounds. The library has to question the veracity of metadata submitted through open and social means: when this metadata combines with library-created metadata, does it result in unintended consequences, such as 'misinformation and conflicting metadata' (Gorzalski 2013, 2)? Other critiques of user-contributed metadata note that it 'lacks structure and reliability,' contains 'ambiguity of meaning,' and is often more 'noise than signal' (Gorman and West quoted in Alemu et al. 2012, 313). This critique insists that a lack of final authority for metadata creation creates superficial, chaotic, untrue, and ultimately untrustworthy data (Alemu et al. 2012). In addition, how much value do these contributions truly add? Some studies of user-contributed metadata have concluded that tagging only added 'low level semantics' of correcting mistakes and adding narrative details (Hooland et al. 2011). In this view, tagging is more about engagement than substantial contribution.

This runs counter to the research of Manzo et al. (2015) who found that, despite these criticisms, user contributions are a net positive:

> Folksonomic metadata, when used in tandem with traditional metadata, increases findability, corrects preventable search failures, and is by and large accurate. Furthermore, the data suggest that given the same tagging conditions, librarians and non-librarians produce a surprisingly similar distribution of useful metadata. (n.p.)

The work of Gross et al. (2015), however, would caution against *replacing* controlled subject access vocabulary with keyword searching and folksonomies; the authors found that approximately 27% of keyword searches in a library catalogue would fail if not for subject controlled vocabularies. This surfaces the notion that the practicalities that our current standardized environment affords are not only for ease of application by library staff, or for machine use. Ultimately, these practicalities also assist users in more reliably and efficiently finding what they are looking for.

But is it possible to reduce library metadata to just practicalities? Some argue that library metadata does not merely facilitate discovery and scholarship but has a larger effect on the greater world of knowledge organization and its social and cultural functions (Andersen and Skouvig 2006). We need to determine how metadata reflects our interaction with cultural heritage in a broader and more indirect sense within society (Hooland et al. 2011). That is, how do our descriptions reflect and shape reality beyond acting as markers for finding particular items?

One clear way to resolve the tension between current library metadata practice and openness is to admit that metadata records are not neutral and fall far short of satisfying the needs of a diverse base of users. Clearly library-produced metadata has limitations, but socially-constructed metadata often lacks syntactic and semantic integrity. The library and its users clearly need both. To establish this, the library needs to recognize its limitations. It needs to be open to other points of view, allow for domain experts to participate, and adopt an iterative approach so that metadata can be changed at any point, destabilizing the idea of a 'perfect,' 'authoritative,' and 'complete' record.

At the same time, the library must remember that metadata is almost always mediated by machine, and even if socially-constructed metadata is preferable, its 'unstructure' does not allow it to be efficiently manipulated and processed by computers. Its inconsistency creates barriers for searching, indexing, and retrieval. With this in mind, the next question surrounds the mechanics of how user-contributed metadata moves from 'comment' to 'canonical.' When users add tags or comments in library catalogues, digital libraries, or institutional repositories, should we even attempt to translate these contributions into standardized language (as often occurs)? Doing so may subvert the idea of open contribution; failure to do so will mean the user contribution does not form part of the metadata 'record of record' and may not be machine-usable.

One approach that allows for a multiplicity of views without sacrificing machine use is linked data. This approach allows anyone to make a machine-understandable statement about anything. It challenges preconceptions that librarians are the final gatekeepers and allows library data to participate in the wider

world of metadata consumption and creation. Examples of social data creation, such as MusicBrainz and Wikidata, show that many users can contribute structured data that can be used as linked data. Indeed, Wikidata has the added benefit of helping to overcome previously prohibitive complexity and expense by circumventing the need to create a bespoke ontology and triple store. In a system like Wikidata, anyone with a Wikimedia account can create structured data immediately usable in a linked data context. The library can then choose to plug into this open and social knowledge creation that happens on a massively larger scale than half-hearted attempts to involve more people in metadata through comments and tagging. The main question is whether the library is ready to embrace the uncontrolled chaos of social knowledge creation. An environment in which the library loses control and power but aligns more with how current users create and consume data is threatening to many in the library, but likely inevitable. The library's failure to reckon with this reality may serve to marginalize it.

Archival and Library Metadata Approaches Contrasted

There are some fundamental differences in the way libraries and archives approach metadata. While archival metadata stretches beyond findability to include structure, context, and authenticity, library metadata generally makes no claim on these and focuses on discovery and content. Special collections and rare book cataloguing are exceptions, as often in these contexts provenance and item-specific information are given. But, depending on focus, which area requires more detailed metadata attention is often arbitrary. A special collection may focus on a particular physical aspect of material (e.g., paper-making technique, binding, etc.) or may pay special attention to a particular person (e.g., illustrator or donor)—in each case, the metadata would emphasize different details. Authors of printed material publish their works with the intention of a broad audience, now and into the future, whereas creators of archives generally do not have posterity in mind. While much of library metadata is explicitly derived from the object in hand—such as title and author— archival metadata is created by an archivist in response to what they can infer from an object. Compare, for example, the description for the book *Who Do You Think You Are?* by Alice Munro, published by MacMillan in 1978, to that of an archival document, a notebook from the Alice Munro archives: 'orange notebook with holograph text. Includes 17 p. poetry, ten untitled fragments and one fragment titled Rapunzel.'

One way to see the differences between archival and library metadata is to look at the difference between databases and narrative. Lev Manovich (cited in Hooland 2011) states that databases and narratives are natural enemies. The database approach focuses on atomized, discrete, unordered, and disconnected metadata fields. The narrative approach creates order and context. This underlines a slight distinction in that library and digital object description is made to be atomized, re-mixed, and matched in whatever order the user finds useful (and this is an impossibility to comprehensively predict), while archival metadata may enable the same re-mixing but wants the context and structure to 'stick' with the metadata and its associated object. The finding aid is a hierarchical document—archival description proceeds from the general to the specific and reflects the arrangement of the documents. Reading an entry in a finding aid without knowing its position within the overall description has the potential to be confusing if not misleading (for example, knowing that a file with a person's name on it is within a grouping of correspondence, as opposed to subject files, and thus contains documents by, rather than about, that person is important to understand). It is not that archival metadata cannot be atomized and re-mixed, as software, such as Access-to-Memory, illustrates, but that the (original) context should always be seen with it in tandem. Archival metadata ideally carries a memory of its context while library metadata usually does not.

As library and archival metadata relate to open social scholarship, the idea that either is truly open is questionable. Descriptions are openly available for use but not necessarily easy to use and accessible. As for open contribution, systems and standards in library and archival metadata act as barriers to widespread participation: although information about objects in a library collection could conceivably be everywhere, the library is not good at integrating them into its metadata records. Archival metadata is probably even less open, in that it requires the content to be described but also must act as the custodian of context and structure. Social contributions are present in both libraries and archives (crowdsourcing transcription, tags, folksonomies, comments, etc.) but are not usually employed at any scale of significance.

One of the greatest differences is the extent to which metadata in either discipline could be considered scholarship. The library view tends to reject this idea. While acknowledging that neutrality in description does not really exist, the library's goal is mostly a description that can satisfy a wide number of uses and contexts that does not attempt to predict what those may be. While acknowledging that research is required to create descriptions, a distinction would be made based on purpose. The purpose of metadata in libraries is to facilitate discovery and scholarship. Archival metadata, on the other hand, is doing more: it is responsible

for enabling discovery of its content by patrons, but also has a duty to be faithful to the structure and context of the archives, and, in turn, the presumption of authenticity of a particular fonds.

Conclusion

An examination of archival and library metadata reveals paths that are variously shared and divergent. The path of archival metadata departs from that of library metadata around the unique documentary nature of the material it describes and the primacy and importance of context. The paths join together around the shared need for international standards and cooperation as well as the desire for consistent metadata than can be easily used by humans and machines. To some extent, this standardization and the barriers to contribution they entail are a shared heritage. While libraries and archives have attempted various ways to allow for open and social contributions to their metadata, these have not generally been effective or employed on a large scale, seemingly undercut by technical barriers, including the lack of broad and unfettered access to descriptive systems and standards. In addition, more open and social metadata creation brings into clearer focus the role and value of professional expertise and, at times, the mistrust in uncontrolled terminology within the profession. More open and social metadata production has the potential to be a disruptor in libraries and archives. While both librarians and archivists have preliminarily explored more social descriptive work, and are in theory open to experimentation, very real practical constraints have limited the realization of radically new approaches to metadata production to date. A profession and practice accustomed to constant change—and the promise and possibilities new technologies enable—will open new paths for metadata within libraries and archives.

Competing Interests

The authors have no competing interests to declare.

References

Alemu, Getaneh, Brett Stevens, Penny Ross, and Jane Chandler. 2012. "The Social Space of Metadata: Perspectives of LIS Academics and Postgraduates on Standards-Based and Socially Constructed Metadata Approaches." *Journal of Library Metadata*, 12(4): 311–44. DOI: https://doi.org/10.1080/19386389.2012.735523

Alemu, Getaneh, Brett Stevens, Penny Ross, and Jane Chandler. 2017. "The Use of a Constructivist Grounded Theory Method to Explore the Role of Socially-Constructed Metadata (Web 2.0) Approaches." *Qualitative and Quantitative Methods in Libraries*, 4(3): 517–40.

Andersen, Jack, and Laura Skouvig. 2006. "Knowledge Organization: A Sociohistorical Analysis and Critique." *The Library Quarterly*, 76(3): 300–322. DOI: https://doi.org/10.1086/511139

Bade, David. 2011. "It's about Time!: Temporal Aspects of Metadata Management in the Work of Isabelle Boydens." *Cataloging & Classification Quarterly*, 49(4): 328–38. DOI: https://doi.org/10.1080/01639374.2011.571096

Buizza, Pino. 2004. "Bibliographic Control and Authority Control from Paris Principles to the Present." *Cataloging & Classification Quarterly*, 38(3–4): 117–33. DOI: https://doi.org/10.1300/J104v38n03_11

Caswell, Michelle. 2017. "Teaching to Dismantle White Supremacy in Archives." *The Library Quarterly*, 87(3): 222–235. DOI: https://doi.org/10.1086/692299

Cook, Terry. 2001. "Fashionable Nonsense or Professional Rebirth: Postmodernism and the Practice of Archives." *Archivaria*, 51: 14–35.

Denton, William. 2007. "FRBR and the History of Cataloging." In: *Understanding FRBR: What It Is and How It Will Affect Our Retrieval*, Arlene G. Taylor (ed.), 35–57. Westport, CT: Libraries Unlimited.

Dooley, Jackie M., and Katherine Luce. 2010. *Taking Our Pulse: The OCLC Research survey of Special Collections and Archives*. Dublin, Ohio: OCLC Research. Accessed August 16, 2018. http://www.oclc.org/research/publications/library/2010/2010-11.pdf.

Drake, J. M. 2016. "RadTech Meets RadArch: Towards a New Principle for Archives and Archival Description." *Medium: On Archivy*. Accessed December 2017. https://medium.com/on-archivy/radtech-meets-radarch-towards-a-new-principle-for-archives-and-archival-description-568f133e4325#.6w1a50egi. Archived at: https://perma.cc/BQJ9-G8T6.

Duff, Wendy, and Kent M. Haworth. 1997. "Advancing Archival Description: A Model for Rationalizing North American Descriptive Standards." *Archives and Manuscripts*, 25(2): 198–199.

Duff, Wendy, and Penka Stoyanova. 1998. "Transforming the Crazy Quilt: Archival Displays from a User's Point of View." *Archivaria*, 45: 44–79.

Duff, Wendy, and Verne Harris. 2002. "Stories and Names: Archival Description as Narrating Records and Constructing Meanings." *Archival Science*, 2(3–4): 263–285. DOI: https://doi.org/10.1007/BF02435625

Gartner, Richard. 2016. *Metadata.* Cham: Springer International Publishing. DOI: https://doi.org/10.1007/978-3-319-40893-4

Gorzalski, Matt. 2013. "Examining User-Created Description in the Archival Profession." *Journal of Archival Organization*, 11(1–2): 1–22. DOI: https://doi.org/10.1080/15332748.2013.866858

Greene, Mark, and Dennis Meissner. 2005. "More Product, Less Process: Revamping Traditional Archival Processing." *The American Archivist*, 68(2): 208–263. DOI: https://doi.org/10.17723/aarc.68.2.c741823776k65863

Gross, Tina, Arlene G. Taylor, and Daniel N. Joudrey. 2015. "Still a Lot to Lose: The Role of Controlled Vocabulary in Keyword Searching." *Cataloging & Classification Quarterly*, 53(1): 1–39. DOI: https://doi.org/10.1080/01639374.2014.917447

Hooland, Seth van, Eva Méndez Rodríguez, and Isabelle Boydens. 2011. "Between Commodification and Engagement: On the Double-Edged Impact of User-Generated Metadata within the Cultural Heritage Sector." *Library Trends*, 59(4): 707–20. DOI: https://doi.org/10.1353/lib.2011.0011

Hyry, Tom, and Michelle Light. 2002. "Colophons and Annotations: New Directions for the Finding Aid." *The American Archivist*, 65(2): 216–230. DOI: https://doi.org/10.17723/aarc.65.2.l3h27j5x8716586q

IFLA. 1997. "Functional Requirements for Bibliographic Records: Final Report." Accessed December 2017. https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf.

InterPARES. 2005. *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project.* Accessed December 2017. http://www.interpares.org/book/index.cfm.

Library of Congress, Network Development and MARC Standards Office. 2018. "MARC 21 Format for Authority Data." Update no. 27, first published in 1999. http://www.loc.gov/marc/authority/.

MacNeil, Heather. 2005. "Picking Our Text: Archival Description, Authenticity, and the Archivist as Editor." *The American Archivist*, 68(2): 264–278. DOI: https://doi.org/10.17723/aarc.68.2.01u65t6435700337

Manzo, Christina, Geoff Kaufman, Sukdith Punjasthitkul, and Mary Flanagan. 2015. "By the People, For the People: Assessing the Value of Crowdsourced, User-Generated Metadata." *Digital Humanities Quarterly*, 9(1): n.p.

Mapping the Stacks. "Home." *Mapping the Stacks: A Guide to Chicago's Hidden Archives.* Accessed December 2017. http://mts.lib.uchicago.edu/.

Millar, Laura. 2004. *Archives: Principles & Practices.* New York: Neal-Schuman.

OCLC. 2016. *"2015–2016 OCLC Annual Report."* Accessed December 2017. http://library.oclc.org/cdm/singleitem/collection/p15003coll7/id/115/rec/36.

Pearce-Moses, Richard. 2005. *A Glossary of Archival and Records Terminology.* Society of American Archivists. Accessed December 2017. http://files.archivists.org/pubs/free/SAA-Glossary-2005.pdf.

Society of American Archivists. 2005. "A*Census." Accessed December 2017. https://www2.archivists.org/initiatives/acensus-archival-census-education-needs-survey-in-the-united-states.

Tanenbaum, Andrew S. 1981. *Computer Networks.* Upper Saddle River, NJ: Prentice-Hall.

Tansey, Eira. 2015. "Archives Without Archivists." *Reconstruction: Studies in Contemporary Cultur*, 16(1): 11.

Taylor, Arlene G. 2004. *The Organization of Information.* Westport, CT: Libraries Unlimited.

Walton, Rachel. 2017. "Looking for Answers: A Usability Study of Online Finding Aid Navigation." *The American Archivist*, 80(1): 30–52. DOI: https://doi.org/10.17723/0360-9081.80.1.30

Yakel, Elizabeth, Seth Shaw, and Polly Reynolds. 2007. "Creating the Next Generation of Archival Finding Aids." *D-Lib Magazine*, 13(5/6): n.p. http://www.dlib.org/dlib/may07/yakel/05yakel.html.