

Knowledge Lost, Knowledge Gained

The Implications of Migrating to Online Archival Descriptive Systems

Daniela Ansovini, Kelli Babcock, Tanis Franco, Jiyun Alex Jung, Karen Suurtamm et Alexandra Wong

Volume 6, numéro 3, 2022

Metadata as Knowledge

URI : <https://id.erudit.org/iderudit/1091361ar>

DOI : <https://doi.org/10.18357/kula.234>

[Aller au sommaire du numéro](#)

Éditeur(s)

University of Victoria Libraries

ISSN

2398-4112 (numérique)

[Découvrir la revue](#)

Citer cet article

Ansovini, D., Babcock, K., Franco, T., Jung, J., Suurtamm, K. & Wong, A. (2022). Knowledge Lost, Knowledge Gained: The Implications of Migrating to Online Archival Descriptive Systems. *KULA*, 6(3), 1–19. <https://doi.org/10.18357/kula.234>

Résumé de l'article

Migrating archival description from paper-based finding aids to structured online data reconfigures the dynamics of archival representation and interactions. This paper considers the knowledge implications of transferring traditional finding aids to Discover Archives, a university-wide implementation of Access to Memory (AtoM) at the University of Toronto. The migration and translation of varied descriptive practices to conform to a single system that is accessible to anyone, anywhere, effectively shifts both where and how users interface with archives and their material. This paper reflects on how different sets of knowledge are reorganized in these shifts. Discover Archives empowers researchers to do independent searches using the full breadth of their domain expertise, seemingly unbound from archival gatekeeping. At the same time, these searches are performed in the absence of archivists' unstructured mediation, where searches benefit from human interaction and the kinds of knowledges that reference staff draw on to handle complex reference questions, especially those from novice archival users. We explore the extent to which that lost knowledge can be drawn back into archival interactions via rich metadata that documents contexts and relationships embedded within Discover Archives and beyond. Internal user experience design (UXD) research on Discover Archives highlights a gap between current online description and habitual user expectations in web search and discovery. To help bridge this gap, we contributed to broader discovery nodes such as linked open "context hubs" like Wikipedia and Wikidata, which can supplement hierarchical description with linked metadata and visualization capabilities. These can reintroduce rhizomatic and serendipitous connections, enabled by archivist, researcher, and larger sets of community knowledges, to the benefit of both the user and the archivist.

RESEARCH ARTICLE

Knowledge Lost, Knowledge Gained: The Implications of Migrating to Online Archival Descriptive Systems

Daniela Ansovini

University of Toronto Archives and Records Management Services

Kelli Babcock

University of Toronto Libraries

Tanis Franco

University of Toronto Scarborough Library

Jiyun Alex Jung

University of Toronto Libraries

Karen Suurtamm

University of Toronto

Alexandra Wong

York University Libraries

Migrating archival description from paper-based finding aids to structured online data reconfigures the dynamics of archival representation and interactions. This paper considers the knowledge implications of transferring traditional finding aids to Discover Archives, a university-wide implementation of Access to Memory (AtoM) at the University of Toronto. The migration and translation of varied descriptive practices to conform to a single system that is accessible to anyone, anywhere, effectively shifts both where and how users interface with archives and their material. This paper reflects on how different sets of knowledge are reorganized in these shifts. Discover Archives empowers researchers to do independent searches using the full breadth of their domain expertise, seemingly unbound from archival gatekeeping. At the same time, these searches are performed in the absence of archivists' unstructured mediation, where searches benefit from human interaction and the kinds of knowledges that reference staff draw on to handle complex reference questions, especially those from novice archival users. We explore the extent to which that lost knowledge can be drawn back into archival interactions via rich metadata that documents contexts and relationships embedded within Discover Archives and beyond. Internal user experience design (UXD) research on Discover Archives highlights a gap between current online description and habitual user expectations in web search and discovery. To help bridge this gap, we contributed to broader discovery nodes such as linked open "context hubs" like Wikipedia and Wikidata, which can supplement hierarchical description with linked metadata and visualization capabilities. These can reintroduce rhizomatic and serendipitous connections, enabled by archivist, researcher, and larger sets of community knowledges, to the benefit of both the user and the archivist.

Keywords: access; archival description; metadata; Wikidata; mediation; user experience

Introduction

Archival arrangement and description—i.e., the representation of archival records as metadata—is an act of knowledge creation performed by archivists, archival staff, and volunteers, who accomplish this work within a specific set of constraints, including both current and legacy archival descriptive standards, institutional policy, archival theory, professional norms, technical systems, and the records themselves. Archival descriptions, created in particular historical, cultural, administrative, and technical contexts, are then carried forward into future contexts and constraints. In 2014, archivists, librarians, and IT/systems workers at the University of Toronto (U of T) began developing an online archival descriptive database using Access to Memory (AtoM), an open-source platform for multi-repository archival description. The site, called [Discover Archives](#), now hosts descriptions for archival collections held in twelve repositories across the university's three campuses. Throughout this work, staff have confronted the logistics of moving legacy finding aids—including unstructured paper and PDF/Word documents and various, idiosyncratic databases—into a single, shared repository of structured data. This work of translating varied descriptive practices to conform to a single system that is accessible to anyone, anywhere, at any time, has significant implications.

Discover Archives has moved U of T into a digital-first archival description workflow and allows search engines to more readily index descriptions for access by the web user. It has also highlighted the limitations of archival descriptive standards when these standards meet modern database architectures and user expectations of online search and discovery experiences. On the one hand, researchers can now search Discover Archives independently, at their own convenience, and those searches benefit from researchers' rich subject knowledge. At the same time, Discover Archives hosts data migrated from finding aids whose structure and content was intended to support researchers' work in the archival reading room, where they benefited from conversations with archival staff and their knowledge of the repository's collections, the records' contexts, and also the limitations of the finding aids themselves—that is, what they could reveal and what they could not. What knowledge is lost in this migration? What knowledge is gained? This paper considers the knowledge implications of the Discover Archives system and service as they reorient traditional relationships between archivists' contextual knowledge of their collections, users' domain knowledge, and the larger sets of professional and community knowledges that may now interact with and feed into Discover Archives. It will also begin to explore how researcher, archivist, and community knowledge might all be brought to bear on archival description, access, and discovery via more robust metadata, linked data, and open knowledge bases like Wikidata.

Our work begins with the understanding that all archival descriptive metadata is an imperfect, mediated representation that emerges from a “fluid, evolving, and socially constructed practice” (Yakel 2003, 2). Archival descriptions aid in search, retrieval, and discovery, acting as surrogates for records and ultimately enabling or constraining researchers' ability to locate the archives they are looking for. Additionally, descriptive metadata is a storying of records that highlights some aspects of an archival fonds or collection while obscuring others; archival description thus constructs meaning and shapes users' understanding of records and the worlds they document. As Wendy Duff and Verne Harris state, “the power to describe is the power to make and remake records and to determine how they will be used and remade in the future. Each story we tell about our records, each description we compile, changes the meaning of the records and re-creates them” (2002, 272). Even once records are located, archival descriptions are read alongside the records in order to understand them because descriptive metadata captures not only record content (what the records can communicate themselves), but also record contexts, structures, and relationships. Archivists follow the principles of provenance and original order and employ hierarchical arrangement and description to preserve and communicate content, context, and structure¹ so that records can make meaning and act as evidence. Quite practically, these principles also help to gain physical and intellectual control over great quantities of archival records, which could never be managed at the item level. Archival records are always described first at the aggregate level (as fonds or collections);² then, depending on resources and the needs of the records, lower-level aggregations (series, subseries, files, and sometimes items) may also be described to provide more granular detail.³ Descriptions of each unit cannot stand on their own but instead are read alongside one another because information at higher levels is not repeated in descriptions for lower levels

¹ The principles of provenance and original order (which together constitute the principle of *respect des fonds*) dictate that archivists keep records created by an individual, family, or organization together as a discrete fonds and that the arrangement of the records should reflect the creator's original arrangement in order to preserve authenticity (Haworth 2001, 12).

² The Discover Archives glossary defines *fonds* as “the highest-level of description, along with ‘Collection’”. The term fonds, originating in French archival practice, can be defined as a body of records that was made and received by a person, family, or organization, public or private, in the conduct of their everyday affairs. These records were accumulated over time and kept for their enduring value as a future reference resource and/or as evidence of the functions and responsibilities of their creator” (Discover Archives).

³ Descriptive approaches such as *Guidelines for Efficient Archival Processing in the University of California Libraries* (Dundon et al. 2020) or “More Product, Less Process: Revamping Traditional Archival Processing” (Greene and Meissner 2005) acknowledge and address the significant resources required for this type of lower-level description and emphasize strategic methods to surface individual records.

(Haworth 2001). Given the complex knowledge embedded in archival descriptive metadata, and the value of that knowledge for search, discovery, and meaning-making, this paper considers the knowledge implications of shifting archival representations from one context (traditional archival description in paper or PDF finding aids, often mediated by an archivist in the reading room) to another (online archival databases, accessed by researchers, often in the absence of archivists). What is the potential for recovering and gaining knowledge in this shift, through both research and a reflective practice?

Discover Archives: The Need for an Online Archival Description Platform

In 2013, archivists at the University of Toronto Archives and Records Management Services (UTARMS) began exploring options for moving their archival finding aids into an archival descriptive database and ultimately selected Access to Memory (AtoM), open-source software developed by Artefactual Systems and widely adopted across Canada and internationally. A February 2014 proposal suggested that “implementation of AtoM at UTARMS would be a significant undertaking, but it would greatly increase control over our holdings, and make our material more manageable, searchable and discoverable for both archivists and researchers” (Suurtamm 2014, 2). At the time, UTARMS’s archival descriptions were distributed across digital and physical spaces, which complicated archival workflows and research. A single online descriptive system held the promise of uniting descriptions, at all levels, through one database, accessible to both staff and researchers around the world. AtoM was chosen as a platform because the software is open source, web-based, and purpose-built for archives; it supports national and international descriptive standards; and it can facilitate interoperability and data sharing.⁴ Its ability to present digital objects alongside existing descriptive metadata and within the context of the larger arrangement of its aggregations was also appealing. But most importantly, AtoM held promise for search, navigation, and discoverability.

UTARMS’s AtoM implementation quickly transformed into a university-wide multi-repository instance. With the aim of creating “the archival parallel to the UTL [University of Toronto Libraries] catalogue,” AtoM promised to be “a significant advance for archives across the university” (Suurtamm 2014, 3). Discover Archives is hosted within U of T Library’s Information Technology Services (ITS) department, and each repository that joins Discover Archives has an archivist on staff to ensure familiarity with AtoM’s built-in archival standards and hierarchical navigation. This shared archival descriptive service is governed by the Discover Archives Steering Committee (DASC), which includes a representative from each repository, along with an ITS digital initiatives librarian, digital initiatives developer, and systems staff who maintain the Discover Archives AtoM instance. At the moment of this paper’s writing, Discover Archives holds more than 115,000 descriptive records, which describe more than 2,200 archival fonds and collections held across twelve repositories. These descriptions are no longer siloed on each archive’s individual website.

Structuring Archival Description as Data

Although the choice of software was important, the fundamental shift happening here was the move from paper-based finding aids to descriptive data stored in a relational database. This shift was initiated by several developments in the archival community, including Encoded Archival Description (EAD) in the 1990s. EAD is an XML data structure standard that supports archival descriptive standards and the complexity of hierarchical archival description. EAD encodes archival metadata in ways that document and preserve content, context, and relationships in machine-readable format (Haworth 2001). At U of T, our move from static finding aids into structured, machine-actionable data came when we adopted AtoM, which improves both functions of archival description: archival management of holdings and archival research. Users can navigate through multi-level descriptions and search at both broad and granular levels. Users can now export, sort, and print box lists. Authority records document relationships amongst creators and between creators and records. AtoM also supports access points (by name, subject, place, and genre) at all levels of description, effectively collating record descriptions in ways unimaginable with PDFs or printed finding aids.

AtoM also generates an XML sitemap that allows pages to be properly indexed for search engine optimization (SEO). Adopting AtoM to build Discover Archives means that descriptions now surface in search engine results for users seeking information on the web. By using a web-based tool like AtoM for archival description, the knowledge contained in descriptions can now also become “part of the web” (Sanderson 2020). Name access points are now links across the system, related descriptions are now structured data stored in AtoM’s MySQL database instead of unstructured text in a Word file, and fonds titles are now URIs to the top-level descriptive record.

⁴ Data sharing with the Ontario provincial descriptive database (Archeion) and the national database (Archives Canada) is especially straightforward, as both platforms also use AtoM.

At first, staff focused on migrating existing descriptive data into AtoM, finding methods and best practices to manually or batch populate Discover Archives with decades of data from paper, PDF, and Word document finding aids. Gregory Wiedeman's "The Historical Hazards of Finding Aids" (2019) helps us understand how this migration work brought the existing limitations of finding aids as a conceptual model of access into the new Discover Archives online space. Wiedeman writes that "when the Internet transformed the possibilities of information access, the limited access goals inherent in finding aids led archivists to prioritize listing materials rather than addressing the broader challenges that users face accessing and using material" (2019, 383). Existing navigation and knowledge limitations of finding aids are also compounded, Wiedeman states, because "implicit connections that were obvious in the reading room became confounding on a screen" (2019, 403).

Initial migration to Discover Archives also continued the document-centric model, an approach that Wiedeman outlines when describing EAD encoding that "prioritized prose over discrete data storage by allowing mixed content and avoided a critical reexamination of dates, extents, or language descriptions" (2019, 400). Some archives only had the resources to migrate top-level (fonds and collection level) descriptions into structured data, and they attached the full PDF finding aid to that description. Lower-level descriptions and detail are then only available in the PDF document, indexed as free text but not structured as hierarchical data in the AtoM database. Still, even the simple act of creating top-level descriptions added an AtoM-enforced layer of structured data on top of the paper finding aid. Figure 1 shows a PDF finding aid from the Trinity College Archives that, in Figure 2, has been entered into Discover Archives through the AtoM web form. Each section of the finding aid then becomes a field, or data points in multiple fields, in AtoM's relational database, lending structure to the flat PDF data.

This first, document-centric migration ensured archives were not losing essential, meticulously crafted data captured in existing finding aids by archival staff who had written descriptions based on their close work with the records and their creators, sometimes decades ago. With structured finding aid data, archivists can now enhance and deepen descriptions and use AtoM's capabilities to build relationships between creators and records. They can also easily export, request, and restructure data to explore how other metadata systems might expose it and build on context. AtoM's front end continues to be document-centric and prioritizes presentation of archival description in a finding aid style, but there are unexplored knowledge implications of this migration in terms of what we can do with finding aid data now that it has been reoriented for online delivery in AtoM. At times, the possibilities seem endless, but these resource-intensive efforts should be informed by our understanding of archival and researcher needs, expectations, and experiences.

Migration: Moving Old Problems Over and Creating New Ones

Researchers benefit when static paper finding aids become online structured data. In her OCLC review of existing research, Jennifer Schaffner (2009) argues that user studies indicate that metadata is key to addressing user needs in online description and discovery. In fact, she finds that for many, "the primary role in discovery is making the collections more visible and [staff] staying out of the way" (Schaffner 2009, 5). At times, Schaffner presents a utopian vision of metadata as a panacea, a vision in which online descriptive systems provide researchers with freedom to search and browse in the absence of the controlling and gatekeeping librarian or archivist. She asserts that "the more users do not need to consult archivists and librarians for searching, the more successful initiatives to improve description and discovery have been" (Schaffner 2009, 5). Researchers can now, in the comfort of their own home or office, search Discover Archives using terms (names, places, events, dates, keywords) that benefit from their own knowledge of their research topic—knowledge that may surpass the archivist's (Yakel and Torres 2003). They can also conduct ongoing, iterative searches without worrying that they are inconveniencing archival staff, and archivists' gatekeeping and control—whether perceived or real—is removed from the site of search.⁵

At the same time, this characterization rests on an oversimplification of the dynamics of archival interactions. Researchers may benefit from archivists' seeming absence online if our descriptive interfaces and metadata seamlessly mesh with user needs and experiences. But archival description is complex, varied, and limited. Archival databases do not behave like Google or even like library catalogues.⁶ Without the archivist there to explain these complexities, variations, and limitations, the researcher-empowered search experience in Discover Archives may actually obscure how a researcher's search results are already mediated by

⁵ For example, in "Liberatory Archives: Towards Belonging and Believing (Part 1)," Jarrett Drake (2016a) identifies silence, solitude, and surveillance as modes of oppressive control identified with penitentiaries that are also the dynamics of archival physical spaces, specifically reading rooms. He points to how these conditions are a continuation of methods of cultural and political exclusion and control and describes the damaging impacts they have on targeted communities.

⁶ In "A 'Major Technological Challenge': Multi-Level Description and Online Archival Databases" (2021), Magnus Berg specifically discusses the structure of relational databases and their incongruous relationship with hierarchical description resulting in the loss of essential context for researchers and opportunities to promote primary source literacy.

Beverley Jones fonds

BEVERLEY JONES FONDS**Dates of creation**

1857- 1935

Extent

13 cm of textual records

Biographical sketch

Beverley Jones was a lawyer who was born 11 June 1839 in Brockville, Ontario, one of three children of Sidney Jones and Susan Ford. Jones was educated at Brockville Grammar School and Upper Canada College before graduating with a BA in 1860 and an MA in 1877, both from Trinity College. He was called to the bar in 1864 and began working in the office of his cousins, Jones Brothers. In 1861 he enlisted in the volunteers at Brockville during the Trent Excitement and served in the Queen's Own Rifles during the Fenian Raid of 1866.

In 1864 Jones joined the Canada Permanent Mortgage Co. as a solicitor and remained in an advisory role there until his death. In 1873 Jones became the bursar of Bishop Strachan School, a private school for girls in Toronto. Jones served as a delegate to the diocesan synod for nearly 50 years as a representative of the congregation of St. George's Church. He also served as secretary of the Canada Law Amendment Association, and was one of the founders of the Industrial Schools Association, serving as treasurer for 30 years. Jones was committed to providing homes for children and established industrial schools for boys and girls, founding the Victoria and Alexandra schools in Mimico and East Toronto. In 1888 Jones drafted the bill known as the Juvenile Offenders Act which provided for a separate trial for juveniles and allowed children under age fourteen to be committed to certain institutions or charitable societies to be taken care of and educated.

Beverley Jones died in Toronto, Ontario in 1934, at age 95.

Scope and content

Fonds consists of notes of lectures by the Provost of the University of Trinity College, delivered in 1859-1860 and notes of lectures on the catechism, as well as Jones's diary as a Trinity College undergraduate from 1857.

Notes

Acquisition history: Unknown

Arrangement: Materials arranged by the archivist

Restrictions on access: None

Terms governing use and reproduction / publication: Public Domain

Related groups of records in different fonds: see file 1-4 in Irving Cameron Fonds (F2067)

Figure 1: A PDF finding aid for the Beverley Jones fonds, held in the Trinity College Archives at U of T.

The screenshot shows the Discover Archives interface for the Beverley Jones fonds (F2097). The header includes the University of Toronto logo and a search bar. The main content area displays the following metadata:

Identity area	
Reference code	CA OTTCA F2097
Title	Beverley Jones fonds
Date(s)	• 1857-1935 (Creation)
Level of description	Fonds
Extent and medium	13 cm of textual records
Context area	
Name of creator	Beverly Jones (1839-1934) Biographical history: Beverly Jones was a lawyer who was born 11 June 1839 in Brockville, Ontario, one of three children of Sidney Jones and Susan Ford. Jones was educated at Brockville Grammar School and Upper Canada College before graduating with a BA in 1860 and an MA in 1... read more
Repository	Trinity College Archives
Immediate source of acquisition or transfer	Unknown
Content and structure area	
Scope and content	The fonds consists of notes of lectures by the provost of the University of Trinity College, delivered in 1859-1860 and notes of lectures on the catechism, including "Questions on Paley's Evidences of Christianity," as well as Jones's diary as a Trinity... read more
System of arrangement	Materials arranged by the archivist.
Conditions of access and use area	
Conditions governing access	Open
Conditions governing reproduction	Public Domain
Finding aids	Finding aid attached
Uploaded finding aid	ottca-f2097.pdf

On the right side of the interface, there are options for 'Finding aid' (Download), 'Clipboard' (Add), 'Explore' (Reports, Browse as list), 'Export' (Dublin Core 1.1 XML), 'Related people and organizations' (Beverly Jones (Creator)), and 'Related genres' (Textual records).

Figure 2: [The finding aid for the Beverley Jones fonds](#) after being transferred to display as-is in Discover Archives.

archivists who have done the work of description and built a system constrained by the limits of their own knowledge and resources as well as the knowledge systems on which their policies, professional norms, standards, and systems are built. In Discover Archives, mediation-via-metadata is more easily effaced by online interfaces, and researchers may be led to believe that archivists have “stayed out of the way,” when instead they have already shaped what is, and is not, possible, retrievable, and knowable.

Historically, reference archivists would help guide researchers to records through the “reference interview, question negotiation, defining and refining search strategies, interpreting finding aids, and providing advice about tools and services” (McCausland 2011, 311). Wendy Duff and Elizabeth Yakel (2017) adopt the term “archival interactions” to characterize archival reference as a site of interaction between four entities: researchers, archivists, archival records, and archival systems. In earlier work, Duff and Yakel, along with Helen Tibbo (2013), interviewed archival reference staff to model the archival reference knowledge (ARK) they bring to this site of interaction. They found that archivists draw on their collection knowledge (including what the repository holds and its contexts), research knowledge (including both methodological and subject knowledge), and interaction knowledge (knowledge of how to interact with people, institutions, and access systems). On site, researchers benefit when archival reference staff’s collection, research, and interaction knowledge complement researchers’ own subject, disciplinary, and methodological knowledge in order to search, retrieve records, and search again, in an iterative fashion.⁷ Although often characterized as a gatekeeper, the archivist also acts as a facilitator whose knowledge of the repository’s collections and varied descriptive practices can help researchers adjust expectations and search strategies according to the limits of imperfect (or simply unfamiliar) metadata and structures. For example, they can clarify why a name or

⁷ Researchers’ methodological and subject knowledge can vary greatly, especially with increased efforts to engage undergraduate students and the general public. Nevertheless, researchers who have done preparatory secondary research, whether students, scholars, or family historians, are likely to have a wealth of subject knowledge they will draw on to develop their search strategies (Yakel and Torres 2003).

keyword search has resulted in zero hits, reorient keyword searches towards provenance-based searches,⁸ and explain that although some collections are described down to the item level, search terms should not be crafted with an expectation of equivalent granularity across the board.

When researchers interact with Discover Archives from outside the archives, both archivists and records are absent from the archival interaction, and we have a new, dyadic interaction between researchers and online archival discovery systems. This dynamic causes a disintermediation of archival interactions, through “the exchange of an archivist mediator for a digital system” (Pugh 2017, 112). The knowledge carried by both records and archivists becomes merely a trace in the online descriptive metadata (McCausland 2011). This leaves us with many questions: how many users leave Discover Archives, frustrated, when their keyword searches receive zero results? How many feel satisfied with a few results when there is actually much more to be found? How many could an archivist have guided to the appropriate records through a simple conversation? There might be ways to digitally reintroduce archivists’ contextual knowledge into Discover Archives metadata, but, in other instances, the system may be unable to replace the kinds of human interactions and relations that get researchers to the records they need.

To develop a deeper understanding of these questions, we can study the experiences of researchers as they navigate Discover Archives and consider how common research questions may be pursued through searches in the database. Two previous user experience studies have been conducted on Discover Archives. The first was completed in 2017 under the direction of Lisa Gayhart, then U of T user experience librarian, with Sori Lee, a graduate student library assistant. The study subjects included first-time undergraduate student users, archivists, and reference librarians, who were asked about their experiences navigating and searching Discover Archives as well as their ability to understand the general purpose of the portal. The resulting report, delivered to DASC, outlined the concerning degree to which students and reference librarians struggled to use the website. Both groups reported difficulties navigating the interface (for example, identifying help and clipboard features). They also had issues understanding the site’s search functionality and search results, in part because of the difficulties interpreting hierarchical archival description. Study participants described being uncertain about distinguishing between the multiple result types for a keyword search represented at varying levels of description as well as authority records for multiple individuals and organizations (Gayhart and Lee 2017). While we anticipated that undergraduate users lacking experience with archives might encounter difficulties, the issues faced by both undergraduates and librarians suggest the extent to which obstacles such as unsupported archival literacy, complex interface navigation, and incongruent expectations might sit in the way of discovery.

In December 2019, we gained further insight from a second user experience testing opportunity, which arose from the User Experience Design (UXD) for GLAM course taught by Olivier St-Cyr at U of T’s Faculty of Information. Led by Tanis Franco, DASC responded to St-Cyr’s call for projects with a series of research questions for the students—in particular, “Do users get an understanding about how archives are structured/described from Discover Archives?” Where the 2017 study had shown us that first-time users of Discover Archives had difficulty understanding archival hierarchy and organization, how could we optimize Discover Archives search functionality and make user experience more intuitive? The UXD study asked six subjects to perform a variety of sample search tasks in Discover Archives, including searches using either a known topic or a known person. The subjects were volunteer users who had little or no prior experience using Discover Archives. Especially for novice users, negative experiences can quickly become discouraging and few will reach out to an archivist if they cannot find what they are looking for.⁹ By studying the volunteer subjects’ processes for finding and interpreting content in Discover Archives, DASC hoped to understand the difficulties faced by users in searching Discover Archives and to implement strategic changes to overcome these barriers.

The UXD project final report summarized four distinct barriers: contact information for archives is difficult to locate; archival vocabulary is opaque to first time users; the search bar’s “search all” functionality is frustrating to users; and the overall front-end display of Discover Archives has excessive information that is not conducive to easy navigation (Braszak et al. 2020). DASC responded by making contact information more visible and standardized across all repositories and by highlighting our glossary page of archival terms. These additions provide opportunities for increased human connections between researchers and archival

⁸ Most U of T archival fonds are organized according to the principle of provenance, which means that records are kept together according to who created them rather than by subject, in order to preserve record context. For this reason, effective archival research looks for records according to who may have created them rather than their topic. Long’s (1989) work on question negotiation in archives describes this kind of mediation well.

⁹ To address this concern, DASC is currently evaluating and revising the help information presented to users on the platform. We aim to more clearly encourage users to contact archivists while also identifying and normalizing some of the reasons they may have difficulty in retrieving relevant results—for example, because not all holdings are represented in the database.

staff, in order to promote the kind of knowledge sharing and support necessary for effective archival interactions. Additionally, DASC hopes to develop approaches that address the information overload that can result from AtoM's default webpage display of archival description and to research how the main search bar can better support users to find what they need.

The UXD studies' findings support Wiedeman's (2019) assertion that document-centric, provenance-based archival discovery systems that are heavily based on hierarchical navigation do not meet users' expectations. At times, user feedback tends to focus on interface and navigation, and testers may take system content for granted, assuming that metadata content and structures are an unchangeable reality. However, archivists, who understand that description emerges from subjective practice, are more likely to consider how issues raised in user studies may also stem from the ways in which metadata affects discovery. Looking at broader research on the preferences of researchers, we can see areas where descriptive practice and arrangement do not fully support the types of discovery that users might expect. For example, Duff and Johnson's (1998) research has shown that historians significantly adapt their searching to accommodate for provenance-based arrangement and the lack of subject access or descriptions of the "aboutness" of records. This finding is reiterated in Schaffner's (2009) work, which considers over thirty years of user studies and highlights the significance of subject-based information to researchers working within special collections and archives.

The reference questions received by archives at U of T confirm that researcher queries are often best supported by metadata capturing "aboutness." Inquiries range in complexity, and archival staff must employ different strategies to address them, from searching known fonds (where the researcher has already identified that the repository has the records of an individual or organization they are studying) to identifying resources that support deep study of historical events and topics. The latter frequently requires thematic or subject searching through multiple sets of sources, formats, and levels of description. It also requires translating reference questions into provenance-based searches (i.e., who would have created these records?), which requires an understanding of whose records are held by the archives in the first place. These types of requests benefit from the reference knowledge of archivists, who can help uncover the contours of a particular question, support archival literacy, and account for the many specificities of a given archival environment and its local practices. For these reasons, complex reference questions quickly challenge online discovery tools and existing metadata, as highlighted in the UXD studies.

Performing subject-based searches in Discover Archives surfaces a variety of issues that affect user experience. For example, a faculty member approached several archival repositories at the University of Toronto to identify materials for student-led digital exhibits on migration and immigrant experiences. In Discover Archives, a keyword search for "[migration](#)" leads to results that include descriptions at the fonds, series, and file level dispersed throughout the list of results. Discover Archives search results include numerous records dealing with movement and relocation of people while also picking up entries on the migration of animals and mountains due to the archives' substantial scientific holdings. While users might expect this broad scope of results from a general keyword search, the information overload and interpretive difficulties noted in the UXD studies are clear. For example, scattered file-level descriptions—some belonging to the same fonds, others not—do little to express the hierarchical relationships between certain results. This display also makes identifying relevant records increasingly complicated because researchers have to sort through a significant volume of results (some of which could be duplicates, given their hierarchical relationships) in order to identify what might be in or out of their research scope.

Aside from the difficulty navigating the display of results, we can also see the limits of keyword searching through aggregate descriptions alone. A Discover Archives search of "migration" surfaces many results that are pertinent to the reference question, including records documenting research on and with specific diasporic communities in Canada as well as studies of immigration policy. However, for keyword searches to retrieve all relevant records, specific terminology must be present in the descriptive metadata. For example, a creator might have migrated and their personal correspondence, journals, or other documentation reflects this experience. A keyword search will miss this material if the descriptions do not use the word *migration* or if the person's fonds is only described at the fonds or series level, where the metadata is unlikely to describe the many events and experiences documented in a single diary.

In responding to the faculty member's reference question, archivists mobilized their collection knowledge to compile a list of people who had immigrated and whose immigration was addressed in some way in the contents of their fonds. Archivists also identified administrative records, such as those of the U of T's International Student Centre—whose link to students who have relocated and the likelihood of students' cultural and social experiences being documented in the records may not come across in fonds-level scope and content notes or file-level titles (e.g., "Council Meetings")—as possible sources. In "A Future without Mediation: Online Access, Archivists, and the Future of Archival Research," McCausland (2011) notes that some scholars have "an ambitious hope that machines will be able to replicate the experience of serendipity

in archival research that derives from the physical proximity of users and references services staff in reading rooms” while others forecast “a continuing role for human interaction in providing reference services” (315). The list of sources that archivists created in response to this reference question on migration benefited from combined sets of knowledge from researchers, who defined the scope and purpose of the research; from reference archivists and their knowledge of collections and descriptive practices; and from the kinds of knowledge embedded in Discover Archives, as keyword searching was able to generate results for some records about which the archivists had little knowledge or recollection. As we ask what possibilities exist for embedding more types of knowledge into our metadata, our understanding of researchers’ desires can help us strategically approach this work in a way that also balances the limitations of descriptive practice, our systems, and the immense scale of the records with which we work.¹⁰ As these various factors interact, no singular approach will fulfill all expectations, but this orientation does help tease out what aspects of discovery might be critical in meeting user needs.

First Steps: Deep Description in Discover Archives

One way to begin bridging the gap between user expectations and Discover Archives’ functionality is to envision and enact the step *after* migration: to spend time redescribing, and more deeply describing, archival holdings within the system, making full use of the rich functionality of AtoM and structured data. Schaffner draws our attention to the important role of metadata in archival discovery, ultimately arguing that “invisibility of archives, manuscripts and special collections may have more to do with the metadata we create than with the interfaces we build. Now that we no longer control discovery, the metadata that we contribute is critical. In so many ways, the metadata *is* the interface” (2009, 4). Given what we know about the knowledge that archivists mobilize to connect users to records in a reference interaction, how much of this archival reference knowledge can be translated into the metadata in Discover Archives?

AtoM supports subject, name, place, and genre access points, which have the potential to support user browsing, collate material in new ways for researchers, and build new intellectual arrangements beyond provenance, thus shifting the ways in which traditional archival description privileges knowledge systems that centre singular notions of creatorship (Drake 2016b; Monks-Leeson 2011; Yeo 2015; Zhang 2012).¹¹ This shift benefits researchers, who encounter descriptions not as static pages but as sets of links and nodes that lead from one set of records to another. One of our instinctive responses to supporting user searches was to consider implementing subject access points, based on the evidence that users enjoy searching and browsing by subject and seeing, at a glance, what a collection is “about” (Duff and Johnson 2002; Stevenson 2008b). Archival implementation of subject heading access has been much slower than in the library world, where cataloguing moved online much earlier (Gabriel 2002). Defining subject terms in a paper finding aid could give researchers a sense of what the records are about, but they could *not* do the work of subject-based collation unless archivists also created subject-based indices. If Discover Archives had a controlled vocabulary of subjects,¹² we could mark certain records as being about “human migration,” for example, whether or not the original descriptions used this terminology. However, our exploration of subject heading implementation in Discover Archives quickly revealed that this work, and its implications for users, was riddled with conceptual and practical issues. Our findings confirmed DASC’s initial decision not to add subject access point metadata. Barriers to consistent application emerge from the lack of an existing standard of practice for applying subject headings to archival description and the varied ways that repositories describe their material in AtoM according to their priorities.¹³ After all, if someone clicks on a link for “human migration,” they—perhaps more so than when doing a keyword search—will expect to retrieve *everything* about

¹⁰ To give a sense of the volume of records that U of T’s repositories hold, UTARMS’s textual holdings exceed eleven linear kilometres (approximately 27,500 banker boxes), with over two hundred thousand photographs and negatives, 200 GB of born-digital material, and tens of thousands of audio recordings, architectural drawings, films, and videos.

¹¹ This shift also responds nicely to calls to expand the notion of provenance (e.g., Douglas 2017) and understand provenance as “now enmeshed with living and ever-fluid frameworks of organizations, communities, individuals, functions, custodians, archivists, and readers, as ‘activators’ of the archive” (Trace 2020).

¹² AtoM supports subject access points and the construction of taxonomies that document connections between broader, narrower, and related terms.

¹³ This would also include unanimous agreement on the subject vocabulary—whether we would implement Library of Congress Subject Headings (LCSH), another common taxonomy, or develop our own. And, of course, these taxonomies already represent and embed particular knowledge systems. For example, many have documented how Library of Congress Subject Headings structure and represent dominant ontologies and epistemologies (Biswas 2018; Howard and Knowlton 2008). However, even given unanimous agreement, consistent application would be required by all archivists working in DASC. Dryden has stated that even with a controlled list of terms for the Provincial Archives of Alberta, the controlled list does “not deal with the differences between content-based and provenance-based indexing or other theoretical aspects of subject access to archival material . . . no guidance in deciding to what level to index (collection, file, item, etc.) . . . nor does it provide any guidance on indexing methodology or detailed instructions on how to index” (1987, 173).

migration. If only half the repositories implemented subject access, the other half's collections would be effectively invisible.

Defining the “aboutness” of record aggregations presents even more complexity (Gabriel 2002; MacNeil 1996). Archives do not typically arrange material by subject and even a single record, such as a letter or diary, can be about dozens of different subjects. When should *human migration* be added as a subject access term in Discover Archives? Should it be added any time the creator of a fonds has themselves migrated, or should it be reserved for instances where we know there are records “about” that migration? When someone has uprooted their life and moved to a new country, which records from that time period are *not* about migration? Is it important to differentiate between records about a creator's migration and records documenting a scholar's research *on* migration, more generally? Is it feasible to add subject terms to the more than 2,200 fonds and collections, or more than 100,000 multi-level descriptions, already described in Discover Archives? And more fundamentally, for university administrative records that only have accession level entries, each accession could conceivably be *about* nearly everything. Mascaro's (2011) study of controlled access headings in EAD finding aids confirms our concerns: she discovered low consistency in application of subject terms and wondered whether LCSH and other subject vocabularies might lack the specificity required to support web retrieval, especially given the tendency to apply terms at higher levels. Given the range of participating archives and their varied collections, subject strengths, users, staffing, priorities, and capacities, our reflection on how to generate thematic or subject access for Discover Archives led to the conclusion that the benefits of subject headings did not justify the immense resources required.

While access points require system-wide, consistent implementation,¹⁴ archivists can already work autonomously, on an ad-hoc basis, to deepen their Discover Archives descriptions by making use of other elements that create structured links and relationships between records, creators, and other entities already described in Discover Archives. For example, archivists can document familial, collegial, and organizational relationships between individuals, families, and organizations. This is done by creating clickable links between authority records but also by using multiple elements to describe the types, natures, and durations of those relationships. AtoM also allows related descriptions to be linked through a controlled field, while providing a free-text field to characterize the nature of that linkage. For example, the Hart House Theatre fonds description makes use of the “Related descriptions” element to connect researchers to records created by related organizations (Hart House), families (the Massey family), and individuals (Jack Gray and Marion Walker) (Hart House Theatre fonds n.d.). These linkages replicate the kinds of conversational links that may have been made in the reading room to direct researchers to more records about the people, organizations, and events they are researching. If these links were more comprehensive throughout the system, Discover Archives could help make the kinds of rhizomatic and serendipitous connections that researchers can make in the reading room when they have access to reference staff and the records themselves (Duff and Haskell 2015). This work also begins to transform archival metadata as actionable knowledge about relationships.

The potential to create access points and document relationships is limited when descriptions are only structured for high-level aggregations of records. Of the 2,280 collections and fonds described in Discover Archives, more than 1,700 are only structured as data at the fonds, collection, or accession level. With time and resources, archival staff could invest in deepening description by creating structured data for series, files, and even some items. For example, the Hart House Theatre fonds is only described at the accession level and, as such, its description lacks the kind of granularity that would reveal that one of those accessions includes an original script by Duncan Campbell Scott called *Joy! Joy! Joy!* (Hart House Theatre fonds n.d.). Searches for Scott or the play's title in Discover Archives do not surface the Hart House Theatre records at all. Ideally, archivists could relate the script to other records by and about Scott held in U of T repositories via his authority record and name access points. But because the theatre's fonds is not described any lower than the accession level, it would be misleading to name Scott as a subject, leading to the impression that a significant part of the accession is about him when the script is just one file amongst sixty-six boxes. Access to this script would require description down to the file level.

Lower-level description also has the potential to surface material about traditionally excluded groups and topics, often obscured when records about people and communities belong to provenance-based fonds organized around a single creator. For example, Dorothy Berry (2018) found that the creation of item-level descriptions, enabled by a digitization project focused on African American materials at the University of

¹⁴ While DASC restricted subject and place access points, it did implement a genre access point, which allows users to limit searches by material type (e.g., textual records, maps). Genre was implemented across all repositories and is added at all levels of description. DASC policy also permits adding name access points at any level of description, which allows archivists to attach authority records to descriptions of archival material about individuals, families, and organizations (rather than just those created by them). However, uneven application of name access points may mislead users who get the impression that, when clicking on a name, they are retrieving all records for that person when they are actually only retrieving instances where a repository has had the capacity to add them.

Minnesota Libraries, led to better description and access precisely because each item now had its own metadata. She argues that aggregate description “privileges majority representation,” making it “less feasible to bring hidden marginalized histories to the forefront” (Berry 2018). Aggregate description limits the kinds of access points and relationships that can be named in the metadata, and even the kinds of keyword searches that will be effective, as long as thousands of records are represented in a single description. Geoffrey Yeo’s (2015) consideration of user needs and technological change also leads him to advocate for lower-level description as a possible solution. Yeo (2015) argues that item-level description with rich metadata facilitates new navigational opportunities and the creation of new arrangements by allowing researchers to collate files and items by multiple facets. These new arrangements are fluid and shifting rather than calcified by a singular archival arrangement, imposed by the archivist, with provenance as the primary organizing principle. Lower-level description will also be necessary if we want to surface, link, and mobilize records and descriptions in sites beyond Discover Archives, as described in the next section.

Digital interfaces have this capacity to enable researcher-led exploration and discovery, but not without significant archival labour and mediation.¹⁵ Looking at the possibilities for archives’ use of linked data, Gracy identifies similar obstacles in applying subject classification and other controlled access points and questions whether the “descriptive richness” of narrative forms of description could be structured as semantic entities (2015, 278). While acknowledging the many challenges of this approach, she ultimately calls for shifts in descriptive standards to more fully support new tools and opportunities for discovery. Similarly, we have considered aspects of description to be one critical factor in understanding the difficulty of meeting user expectations and question how and where linked data could offer new approaches in creating more dynamic, inclusive, and fruitful systems. Can linked data help to preserve and represent complex relationships between metadata sets and to leverage not only archivist and researcher knowledge but also broader sets of community knowledge about the worlds documented in our records?

Linked Data and Future Opportunities for Metadata Mediation

Our group’s reflections on subject access in Discover Archives led us to conclude that there is enormous complexity in representing the aboutness of records. Applying subjects requires a difficult distillation that is incomplete to present to users in the form of subject search facets. Inconsistent, unreliable, or broad application could also potentially undermine the value of Discover Archives for researchers and “could exacerbate rather than mitigate the problem of an overabundance of material” (Lippincott 2021, 59–60). Affirming to users that they could easily contact a human archivist for help, while also deepening description by documenting other relationships via metadata, seemed like a far better return on time invested than collectively struggling to resolve the complexities of subject terms and then identifying and applying them. We thus redirected our discussion of subject-oriented discovery to reflect on the new opportunities presented by Discover Archives’ structured finding aid metadata.

First, we considered the user’s journey through the web to land on the Discover Archives website. From the perspective of the user, discovery consists of any number of steps between the start of information search and the resulting information retrieval. This process is not necessarily linear, and the layers of mediation experienced during those steps will affect what information users find. The user may make their journey through a variety of nodes and across a range of tools, including but not limited to search engines and Wikipedia, which are often more accessible to them by virtue of habit and exposure. Search engines and Wikipedia are referring a large number of users to our archival discovery system, and many users prefer to search broadly across the web when beginning research (DeRidder 2008; Lippincott 2021). Here, linked open databases like Wikidata and DBpedia feed into search algorithms and knowledge graphs.¹⁶ The extent to which archival holdings are linked from these nodes, which researchers pass through during their web-based discovery journey, helps direct them to Discover Archives, where they can then deepen their search or contact an archivist.

¹⁵ As Berry (2018) notes, “this sort of detailed work is not financially viable or efficient enough to fit into the workflow of generally already over-taxed processing departments, but in a project funded situation can greatly assist in recognition and accessibility.” Although the university’s libraries and archives lack the human and financial resources to support the tremendous labour that would be required by any systematic deep description projects, some of this work can also be done on an ongoing, ad-hoc basis. Allain and Robichaud (2017) also offer more insight into this issue. They outline how the many complexities of archival description work have led to the current situation of archives not being able to “just script it” when putting archival description data online in the same way that, for example, can be done for standardized bibliographic data in library discovery systems.

¹⁶ DBpedia extracts structured content from Wikipedia and is a central hub in the web of data. Wikidata is a similar hub that is also integrated by design with Wikipedia, but is neutral on data source (it is openly editable). Both databases can be queried for contextual exploration using SPARQL, a query language.

In recognition of this, archivists added external links in Wikipedia to lead users to finding aids. These contributions were sometimes removed by Wikipedia editors, rightly cautious that a user with no understanding of archival material will be confused to find themselves on a descriptive page without direct, web-based access to materials. More recently, we built on past work through the U of T Wikipedian-in-Residence program (2018–19) to sensitively and usefully contribute archival information related to the subject of an article. These contributions include Wikipedia articles for scholar-activists Roxana Ng and Rodney Bobiwash, which use finding aid biographies as sources, as well as a template ([Template:Archival records](#)) and its component help page ([Help:Archival material](#)), which locate archival holdings and fold in archival literacy for the user. Appropriately contextualized external links from Wikipedia can direct users to archival holdings through connective tissue that bridges the jump from tertiary source to primary records (see Figure 3).¹⁷

As standalone discovery infrastructure, linked open databases constitute networks of information that can be explored rhizomatically, through and around different nodes, without requiring archivists to build these nodes independently. A proof of concept for this exists in the “Pilote d’interopérabilité pour les autorités archivistiques françaises” (PIAAF) pilot project from the International Council on Archives (ICA) Expert Group on Archival Description (EGAD), based on the Records in Context Conceptual Model (Clavard 2018). In 2019, DASC was inspired to cultivate the possibilities of linked data after a thought-provoking Association of Canadian Archivists conference on the topic (McLellan et al. 2019; Wong and Babcock 2021). Among our takeaways were that name access points found in archival description are ripe for conversion into linked data entities; linked data offers users the ability to navigate context beyond the constraints of archival theory; descriptions for the open web would impact mediation and what an archivist chooses to describe; and query visualizations could be offered to users for common reading room questions. Exploring these opportunities through Wikidata now can inform eventual assessment and adoption of standard linked data ontologies for archival description, such as Records in Context – Ontology (RiC-O) once the final version is published (ICA Expert Group on Archival Description 2021).

In the months following, we reviewed other institutions’ Wikidata projects and made decisions regarding our own Wikidata ontologies and workflows (Cohen-Palacios 2019; Grguric et al. 2019). We created data models to link Wikidata’s name entities to top-level descriptions in Discover Archives by their holding institution (“Wikidata:WikiProject University of Toronto Libraries/Discover Archives - Wikidata” n.d.). Throughout 2020 and 2021, we created hundreds of new items for people and organizations based on Discover Archives authority records and linked all pertinent holdings using the “archives at” property, as seen in Figure 4.

Through this work, we gained several new entities and their properties as community-generated access points contributed by unaffiliated editors.¹⁸ We also noticed multiple opportunities for metadata mediation, as McLellan et al. (2019) had predicted. In the example for Margaret Atwood, we created a semantic triple allowing knowledge to be represented in both a human and machine-readable format of subject – predicate – object: Margaret Atwood – archives at – Thomas Fisher Rare Book Library.¹⁹ This triple is surrounded by other triples, developed by other Wikidata contributors, that all add relational context to the Fisher’s archival holdings for Margaret Atwood. These triples include entities for each of Atwood’s novels, publication dates, names of her family members, and her literary influences (e.g., Margaret Atwood – author – *The Handmaid’s Tale*).

Visualizations like Figure 5 can enable a new form of online mediation for archives—a web of access points that can be mobilized by users as sources for additional keywords or context alongside Atwood’s finding aids and records. The simple addition of one “archives at” triple, linking the records to an item in Wikidata, allows for the kinds of connections that may have been made in the reading room, where archivists use their contextual knowledge to suggest to researchers records of related people, important dates, and notable publications. Wikidata can provide machine-actionable, community-generated access points that allow users to perform exploratory searches to visualize, search, and explore relationships to other entities.²⁰ In this way, our intentional participation in the linked open web of data can supplement our

¹⁷ An overview of Wikidata in Wikipedia infoboxes and [Template:Archival records](#) can be found in a recording of the LD4 Wikidata Affinity Group meeting on May 5, 2020 (Jung 2020).

¹⁸ We are reminded that Wikidata was born of a vision for a semantic Wikipedia (Erleben et al. 2014; Krötzsch et al. 2006) and that Wikipedia manifests the idea that “no one knows everything but everyone knows something.” This phrasing, found time and again in Wikipedia user essays and talk pages, is a version of the age-old parable “blind men and an elephant.”

¹⁹ In other words, Thomas Fisher Rare Book Library (the institution home of the Margaret Atwood fonds) is linked to her Wikidata item using the “archives at” property. The property allows a qualifier through which the description URL can be appended.

²⁰ Speaking on user interfaces that allow exploratory search, Dan Scott (2020) explained at the Access 2020 conference that exploratory search involves moving “from a familiar domain through an unfamiliar domain, using the information we access to overcome conceptual and knowledge gaps. In developing exploratory search interfaces, then, our goal is to help people bridge those gaps, and construct (and deconstruct) the sense they make of their worlds.”

External links [edit]

- [Roxana Ng fonds at the University of Toronto Archives](#)
- [Homeworkers Association records at the University of Toronto Archives](#)

Archives at	
Location	University of Toronto Archives & Records Management Services
Identifiers	B2014-0005 ✎
Source	Roxana Ng fonds ✎
How to use archival material	

Figure 3: External links displayed in Wikipedia. Template values can be entered manually in Wikipedia or pulled dynamically from Wikidata (indicated by blue pencil icon). Manual entry overrides Wikidata-sourced values.

archives at

Thomas Fisher Rare Book Library ✎ edit

described at URL <https://discoverarchives.library.utoronto.ca/index.php/margaret-atwood-papers>

▶ 1 reference

+ add value

Figure 4: Example of “archives at” property on [Margaret Atwood’s Wikidata item](#).

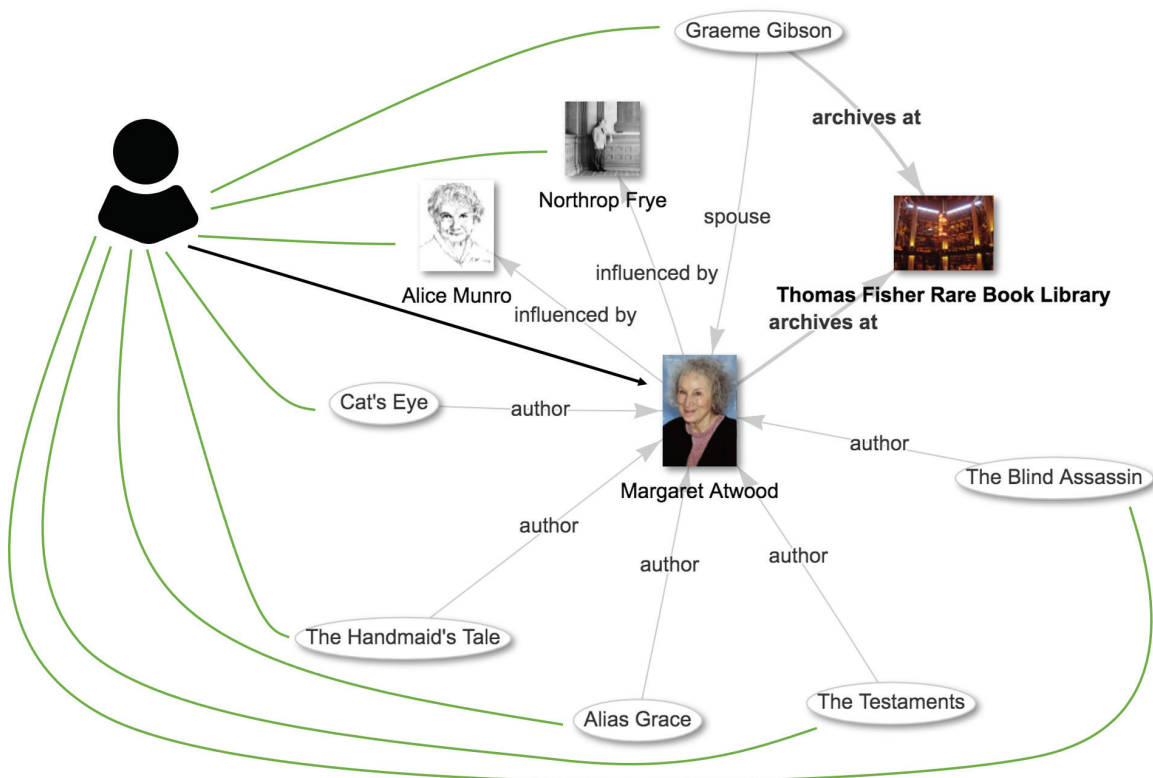


Figure 5: This [SPARQL query](#) of select items in Wikidata that directly link to or from the Margaret Atwood item visualizes community-generated access points (green lines) to Margaret Atwood and to her archival holdings at the Thomas Fisher Rare Book Library, in addition to the regular direct access that Discover Archives provides (black arrow).

hierarchical and narrative archival discovery systems with participatory mediation. Since anyone with a Wikidata account can build onto the Margaret Atwood Wikidata item, the community can add more context and access points organically over time, allowing for multiple instances of meaning construction.²¹

Our migration to Discover Archives has signaled the need to explore new ways of enhancing description, access, and reference practices using these new opportunities for online mediation. Google Analytics reveals that our users are taking diverse journeys that refer them to our websites, and UXD research reveals that they want more entryways into the records when navigating our online archival discovery systems. Rather than posit a singular solution to fulfill user needs, we can strengthen the structured finding aid data in Discover Archives through deeper description and use existing structured data to build new forms of online mediation, all while surfacing this mediation in the user's web journeys and making sure that researchers can easily connect with an archivist from our descriptive databases. The goal of creating linked data out of structured finding aid data is not to *replace* finding aids, but to better understand how archivists can help users make connections and meaning between records when they are researching online—possibly an even more important goal than struggling to generate finite subject terms that may only provide very linear, surface-level access within our own closed discovery systems.

Our initial investigations into Wikidata were expedited by the existence of Discover Archives. Our linked data exploration is an appendage to the structured data that Discover Archives has provided to our finding aids and is the result of many hours of labour by archival staff who have moved descriptions into the database. Online mediation in the absence of an archivist still relies on the work of archivists and requires trained archival staff to create meaning through archival description and context building. It is only because of archivists' labour that linked data offers the opportunity to avoid “the historical hazards of finding aids” that Wiedeman (2019, 381) identified. Archivists still hold the context knowledge of the fonds and its relationships to the larger collection, expertise gained only from working closely with archival creators and through processing time with the records. The mediation opportunities we can explore now with linked data simply offer archivists and the creators of archival discovery platforms an opportunity to reformat that knowledge into a new structure and invite collaborative context building through tools such as Wikidata or others—an additional storying of archival description metadata and its entities through “parallel description” (Babcock et al. 2021, 105). In moving from traditional archival description to presenting entities and relationships online, archivists have to consult users to better understand how linked data metadata can help mediate their research needs.

Briefly, there are other mediation opportunities to consider in moving archival description to linked data. Archival descriptions can be fraught with bias, sanitization, or missing context and, in some cases, descriptions may be in need of recontextualization. For example, Douglas et al. (2018) propose linked data as a way for archives to begin working on recontextualizing descriptions as part of work towards decolonisation. Although they do not name Wikidata, their proposal is preceded by metadata decolonisation work put into practice in Wikidata by Indigenous communities and libraries (Allison-Cassin and Scott 2018). Wikidata as a community-based tool for redescription is an area worth further research.

In Wikidata's open technology community, metadata can be freely created by community groups and is thus collectively mediated. Seeman and Dean point out the tension embedded in opening up library and archival metadata for community contribution since archival description “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” (2019, 9). However, they note earlier that “the main question is whether the library [or archive] is ready to embrace the uncontrolled chaos of social knowledge creation. An environment in which the library [or archive] loses control and power but aligns more with how current users create and consume data is threatening to many in the library [or archive], but likely inevitable” (Seeman and Dean 2019, 8). One possible opportunity is to engage with researchers whose domain expertise could enhance items in Wikidata, which could then be reviewed by archivists to augment Discover Archives. For example, Figure 6 shows how a researcher or archivist could create a Wikidata item for the original *Joy! Joy! Joy!* manuscript, link it to its creator, Duncan Campbell Scott, and add an “archives at” from the script to its Discover Archives holdings. Though the Discover Archives system does not encourage such an incomplete description of one file and inconsistent linkages of lower-level items, Wikidata, as a platform in constant evolution, complements the Discover Archives discovery by allowing this parallel description and additional context. Researchers and archivists can add properties or items in Wikidata to continue contextualizing the play—for example, linking

²¹The work of harnessing Wikidata for participatory mediation that serves just description, however, is not to be taken lightly. Wikidata's openness can be used to generate data that is harmful or violent just as much as it can provide opportunity for reparative description. Mizota (2021) describes an instance of a Wikidata item being vandalized with racial slurs as well as the harm caused from erasure when terms are absent.

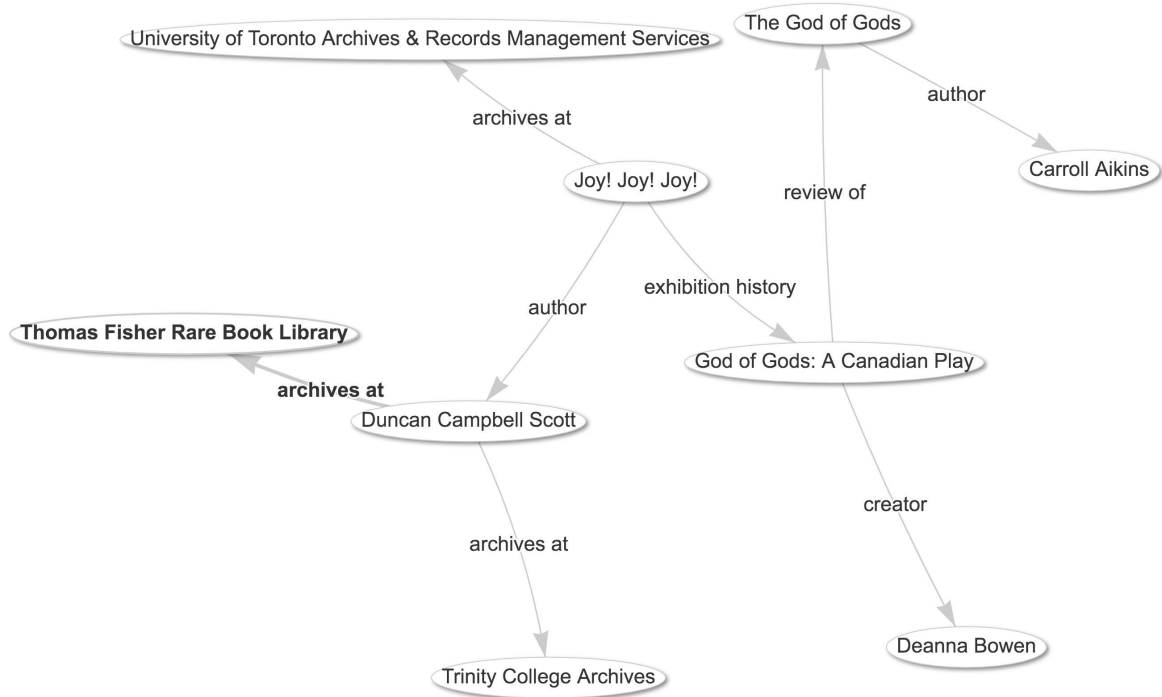


Figure 6: Sample query of context around a record using the *Joy! Joy! Joy!* manuscript from the Hart House Theatre accession at UTARMS (<https://w.wiki/3ZD5>).

its inclusion in Deanna Bowen’s exhibit *God of Gods: A Canadian Play* (Art Museum at the University of Toronto n.d.), which included one of UTARMS’s photographs of a Hart House Theatre production of *Joy! Joy! Joy!*—and put it into dialogue with other material held across many archives.

It behooves us to reckon with these opportunities and constraints sooner rather than later. Though open, Wikidata still tends to replicate dominant ontologies through skewed participation and gatekeeping norms. Libraries and archives are just starting to develop best practices in Wikidata. We must notice who is at the table in developing these practices and consider the varied needs of archival users. Sadler and Bourg (2015) consider how to consciously build discovery systems and access points that can avoid the “consensus-based relevancy” of anything-but-neutral search engines like Google and the historic bias woven into many archives and library classification systems. They propose that we consider using feminist human computer interaction qualities in thinking about online discovery, including plurality, self-disclosure, participation, ecology, advocacy, and embodiment. Further study is needed to assess these qualities when creating linked data for archival descriptions, but Sadler and Bourg’s (2015) work offers one model to address the inherent bias in linked data metadata work.

Conclusion

Our analysis of Discover Archives, and the linked data opportunities it inspires, reveals how complex sets of knowledge interact to make archival search and discovery possible. Access is not limited to online archival description; it also benefits from modes of archival mediation, human-to-human interaction, and knowledge sharing between record creators and communities, archivists, and researchers. Research on how users navigate Discover Archives compels us to explore how we might transfer some of this knowledge through metadata itself: first, by embedding archival mediation into our interfaces and descriptions, and second, by layering knowledge held both within and outside archival contexts. Discover Archives provides immediate access to archival descriptions across U of T, puts us on a path toward better search engine-based discovery, and enables us to map our holdings into the vast, decentralized web of openly edited and referenced data. Deepening our understanding of how users navigate online tools can help us develop metadata that leverages multiple knowledges to support the rhizomatic and dynamic forms of discovery that happen in the reading room. As a group of repositories with differing resources, priorities, and practices, how do we initiate these shifts in ways that are realistic, sustainable, and responsive to changing needs from a variety of actors? At U of T, this may include identifying and prioritizing collections for deeper description

in Discover Archives, which would enhance our capacity to create meaningful contextual relationships in Wikidata and elsewhere. We could also invite researchers and communities with subject expertise to guide and inform this work, acknowledging their similar investment in enhancing our resources with additional context, nuance, and meaning. Any approach needs to be acutely aware of what knowledge, and whose knowledge, is lost and gained in these shifts, and to create opportunities to embed that knowledge within our descriptions.

Acknowledgments

The work and thinking documented in this paper is indebted to the ideas, efforts, and labour of many others. The authors want to first thank the editors of the special issue, Stacy Allison-Cassin and Dean Seeman, and our reviewers for their helpful feedback. Thank you to our colleagues in various community groups as well as U of T libraries and archives staff involved in Discover Archives, the Wikipedian-in-Residence program, and the Wikidata “archives at” work.

Competing Interests

The authors declare that they have no competing interests.

References

- Allain, Sara, and Danielle Robichaud. 2017. “No, We Can’t Just Script It: And Other Refrains from (Tired) Archival Data Migrators.” Presentation at the Access Conference 2017, Saskatoon, SK, September 28, 2017. YouTube video, 17:20. https://www.youtube.com/watch?v=SHoua98ZMfy&list=PLomHagvStAaDzXullxohONtPPcD_T2AO4&index=8.
- Allison-Cassin, Stacy, and Dan Scott. 2018. “Wikidata: A Platform for Your Library’s Linked Open Data.” *Code4Lib Journal* 40. <https://journal.code4lib.org/articles/13424>.
- Art Museum at the University of Toronto. n.d. “God of Gods: A Canadian Play.” Accessed May 6, 2022. <https://artmuseum.utoronto.ca/exhibition/deanna-bowen-the-god-of-gods/>.
- Babcock, Kelli, Regine Heberlein, Anna Björnsson McCormick, Elizabeth Russey Roke, Greta Kuriger Suiter, and Ruth Kitchin Tillman. 2021. “The Power of Parallel Description: Wikidata and Archival Discovery.” In *The Lighting the Way Handbook: Case Studies, Guidelines, and Emergent Futures for Archival Discovery and Delivery*, edited by Mark A. Matienzo and Dinah Handel, 99–113. Stanford, CA: Stanford University Libraries. <https://doi.org/10.25740/gg453cv6438>.
- Berg, Magnus. 2021. “A ‘Major Technological Challenge’: Multi-level Description and Online Archival Databases.” *Emerging Library & Information Perspectives* 4 (1): 62–87. <https://doi.org/10.5206/elip.v4i1.12529>.
- Berry, Dorothy. 2018. “Digitizing and Enhancing Description Across Collections to Make African American Materials More Discoverable on Umbra Search African American History.” *The Design for Diversity Learning Toolkit*. Northeastern University Library. <https://des4div.library.northeastern.edu/digitizing-and-enhancing-description-across-collections-to-make-african-american-materials-more-discoverable-on-umbra-search-african-american-history/>.
- Biswas, Paromita. 2018. “Rooted in the Past: Use of ‘East Indians’ in Library of Congress Subject Headings.” *Cataloguing & Classification Quarterly* 56 (1): 1–18. <https://doi.org/10.1080/01639374.2017.1386253>.
- Braszak, Lucas, Olivia Chlebicki, Andrew Edmonds, Jamie Lee Morin, and Katharine Taylor. 2020. “Project: Research and Interpretation [U of T Discover Archives UXD Project].” Unpublished report.
- Clavaud, Florence. 2018. “Semantizing and Visualising Archival Metadata: The PIAAF French Prototype Online.” International Council on Archives. <https://www.ica.org/en/semantizing-and-visualising-archival-metadata-the-piaaf-french-prototype-online>.
- Cohen-Palacios, Katrina. 2019. “Wikidata and Archivists.” Presentation at the Archives Association of Ontario (AAO) Institutional Issues Forum, Toronto, ON, October 24, 2019. <http://hdl.handle.net/10315/36898>.
- DeRidder, Jody L. 2008. “Googlizing a Digital Library.” *Code4Lib Journal* 2. <https://journal.code4lib.org/articles/43>.
- Discover Archives. “Glossary.” <https://discoverarchives.library.utoronto.ca/index.php/glossary>.
- Douglas, Jennifer. 2017. “Origins and Beyond: The Ongoing Evolution of Archival Ideas about Provenance.” In *Currents of Archival Thinking*, 2nd ed., edited by Heather MacNeil and Terry Eastwood, 25–52. Santa Barbara, CA: Libraries Unlimited.
- Douglas, Jennifer, Greg Bak, Evelyn McLellan, Seth van Hooland, and Raymond Frogner. 2018. “Decolonizing Archival Description: Can Linked Data Help?” *Proceedings of the Association for Information Science and Technology* 55 (1): 669–72. <https://doi.org/10.1002/pra2.2018.14505501077>.

- Drake, Jarrett M. 2016a. "Liberatory Archives: Towards Belonging and Believing (Part 1)." On Archivy. <https://medium.com/on-archivy/liberatory-archives-towards-belonging-and-believing-part-1-d26aaeb0edd1>.
- Drake, Jarrett M. 2016b. "RadTech Meets RadArch: Towards a New Principle for Archives and Archival Description." On Archivy. <https://medium.com/on-archivy/radtech-meets-radarch-towards-a-new-principle-for-archives-and-archival-description-568f133e4325>.
- Dryden, Jean E. 1987. "Subject Headings: The PAASH Experience." *Archivaria* 24: 173–80.
- Duff, Wendy M., and Verne Harris. 2002. "Stories and Names: Archival Description as Narrating Records and Constructing Meanings." *Archival Science* 2: 263–85. <https://doi.org/10.1007/BF02435625>.
- Duff, Wendy M., and Jessica Haskell. 2015. "New Uses for Old Records: A Rhizomatic Approach to Archival Access." *The American Archivist* 78 (1): 38–58. <https://doi.org/10.17723/0360-9081.78.1.38>.
- Duff, Wendy M., and Catherine A. Johnson. 2002. "Accidentally Found on Purpose: Information-Seeking Behavior of Historians in Archives." *Library Quarterly* 72 (4): 472–96. <https://www.jstor.org/stable/40039793>.
- Duff, Wendy, and Penka Stoyanova. 1998. "Transforming the Crazy Quilt: Archival Displays from a User's Point of View." *Archivaria* 45: 44–79. <https://archivaria.ca/index.php/archivaria/article/view/12224>.
- Duff, Wendy, and Elizabeth Yakel. 2017. "Archival Interaction." In *Currents of Archival Thinking*, 2nd ed., edited by Heather MacNeil and Terry Eastwood, 193–224. Santa Barbara, CA: Libraries Unlimited.
- Duff, Wendy M., Elizabeth Yakel, and Helen Tibbo. 2013. "Archival Reference Knowledge." *American Archivist* 76 (1): 68–94. <https://www.jstor.org/stable/43489650>.
- Dundon, Kate, Laurel McPhee, Elvia Arroyo-Ramirez, Jolene Beiser, Courtney Dean, Audra Eagle Yun, Jasmine Jones, et al. 2020. *Guidelines for Efficient Archival Processing in the University of California Libraries (Version 4)*. UCLA Library. <https://escholarship.org/uc/item/4b81g01z>.
- Erleben, Fredo, Michael Günther, Markus Krötzsch, Julian Mendez, and Denny Vrandečić. 2014. "Introducing Wikidata to the Linked Data Web." In *The Semantic Web – ISWC 2014*, edited by Peter Mika, Tania Tudorache, Abraham Bernstein, Chris Welty, Craig Knoblock, Denny Vrandečić, Paul Groth, Natasha Noy, Krzysztof Janowicz, and Carole Goble, 50–65. Cham: Springer International. https://doi.org/10.1007/978-3-319-11964-9_4.
- Gabriel, Claire. 2002. "Subject Access to Archives and Manuscript Collections: An Historical Overview." *Journal of Archival Organization* 1 (4): 53–63. https://doi.org/10.1300/J201v01n04_04.
- Gayhart, Lisa, and Sori Lee. 2017. "Discover Archives Usability Interviews: Findings and Recommendations." Unpublished report.
- Godby, Jean, Karen Smith-Yoshimura, Bruce Washburn, Kalan Knudson Davis, Karen Detling, Christine Fernsebner Eslao, Steven Folsom, et al. 2020. *Creating Library Linked Data with Wikibase: Lessons Learned from Project Passage*. OCLC Research. <https://doi.org/10.25333/faq3-ax08>.
- Gracy, Karen F. 2015. "Archival Description and Linked Data: A Preliminary Study of Opportunities and Implementation Challenges." *Archival Science* 15: 239–94. <https://doi.org/10.1007/s10502-014-9216-2>.
- Greene, Mark A., and Dennis Meissner. 2005. "More Product, Less Process: Revamping Traditional Archival Processing." *The American Archivist* 68 (2): 208–63. <https://doi.org/10.17723/aarc.68.2.c741823776k65863>.
- Grguric, Ekatarina, Frédéric Giuliano, Anna Dysert, and Rachel Black. 2019. "From Boxes to AtoM to Wikidata." Presentation at the Access Conference, Edmonton, AB, September 30, 2019. <https://accessproceedings.ca/index.php/access/article/view/86>.
- Hart House Theatre fonds. n.d. University of Toronto Archives, Toronto, Ontario. <https://discoverarchives.library.utoronto.ca/index.php/hart-house-theatre-fonds>.
- Jung, Alex. 2020. "Wikidata and Wikipedia Infoboxes." Presentation at the LD4 Wikidata Affinity Group meeting, May 5, 2020. https://docs.google.com/document/d/1xSuAWN01FVfOlvSML2_7Hzfn0PVoGH-gARyGB1tcL6No/edit.
- Haworth, Kent M. "Archival Description: Content and Context in Search of Structure." *Journal of Internet Cataloging* 4 (3–4): 7–26. https://doi.org/10.1300/J141v04n03_02.
- Howard, Sara A., and Steven A. Knowlton. 2018. "Browsing through Bias: The Library of Congress Classification and Subject Headings for African American Studies and LGBTQIA Studies." *Library Trends* 67 (1): 74–88. <http://doi.org/10.1353/lib.2018.0026>.
- International Council on Archives (ICA) Expert Group on Archival Description. 2021. "Records in Contexts – Ontology." International Council on Archives. <https://www.ica.org/en/records-in-contexts-ontology>.
- Lippincott, Sarah. 2021. *Mapping the Current Landscape of Research Library Engagement with Emerging Technologies in Research and Learning: Final Report*. Edited by Mary Lee Kennedy, Clifford Lynch, and Scout Calvert. Association of Research Libraries, Born-Digital, Coalition for Networked Information, and EDUCAUSE. <https://doi.org/10.29242/report.emergingtech2020.landscape>.

- Long, Linda J. 1989. "Question Negotiation in the Archival Setting: The Use of Interpersonal Communication Techniques in the Reference Interview." *The American Archivist* 52 (1): 40–50. <https://www.jstor.org/stable/40293311>.
- MacNeil, Heather. 1996. "Subject Access to Archival Fonds: Balancing Provenance and Pertinence." *Fontes Artis Musicae* 43 (3): 242–58. <https://www.jstor.org/stable/23508211>.
- McCausland, Sigrid. 2011. "A Future Without Mediation? Online Access, Archivists, and the Future of Archival Research." *Australian Academic & Research Libraries* 42 (4): 309–19. <https://doi.org/10.1080/00048623.2011.10722243>.
- McLellan, Evelyn, Anna Dysert, Krista Jamieson, and Katherine Timms. 2019. "Is Linked Data (LD) the Future of Archival Description?" Paper presented at the Association of Canadian Archivists Conference, Toronto, ON, June 6–8, 2019. https://www.archivists.ca/resources/Documents/Conference%20Material/Past%20Conference%20Documents/Conference%20Programs/20190605_Archival%20Origins%20Toronto%20V%203.1.pdf. Archived at: <https://perma.cc/Q9FA-UZB7>.
- Mizota, Sharon. 2021. "Change Is Good: Navigating Wikidata as a Controlled Descriptive Vocabulary." *Descriptive Notes* (blog). March 30, 2021. <https://saadescription.wordpress.com/2021/03/30/change-is-good-navigating-wikidata-as-a-controlled-descriptive-vocabulary/>. Archived at: <https://perma.cc/5KP3-EP2C>.
- Monks-Leeson, Emily. 2011. "Archives on the Internet: Representing Contexts and Provenance from Repository to Website." *The American Archivist* 74 (1): 38–57. <https://doi.org/10.17723/aarc.74.1.h386n333653kr83u>.
- Pugh, Joseph Jonathan. 2017. "Information Journeys in Digital Archives." PhD diss., University of York. <https://etheses.whiterose.ac.uk/20663/1/Proofed%20Corrected%20thesis.pdf>.
- Sadler, Bess, and Chris Bourg. 2015. "Feminism and the Future of Library Discovery." *Code4Lib Journal* 28. <https://journal.code4lib.org/articles/10425>.
- Sanderson, Robert. 2020. "The Importance of Being LOUD." Report presented at LODLAM Summit, Los Angeles, CA, February 3–4, 2020. <https://www.slideshare.net/azaro42/the-importance-of-being-loud>.
- Schaffner, Jennifer. 2009. "The Metadata *Is* the Interface: Better Description for Better Discovery of Archives and Special Collections, Synthesized from User Studies." OCLC Research. <https://library.oclc.org/digital/collection/p267701coll27/id/444/>.
- Scott, Dan. 2020. "LINCS: Exploratory Search." Paper presented at the Access Conference, virtual, October 19–23, 2020. <https://drive.google.com/file/d/1AhXOQK7jScC6LSxI-ATT6FgXNPU4DbHG/view>.
- Seeman, Dean, and Heather Dean. 2019. "Open Social Knowledge Creation and Library and Archival Metadata." *KULA: Knowledge Creation, Dissemination, and Preservation Studies* 3: 13. <https://doi.org/10.5334/kula.51>.
- Stevenson, Jane. 2008a. "The Online Archivist: A Positive Approach to the Digital Age." In *What Are Archives? Cultural and Theoretical Perspectives: A Reader*, edited by Louise Craven, 89–108. Aldershot: Ashgate.
- Stevenson, Jane. 2008b. "'What Happens if I Click on This?' Experiences of the Archives Hub." *Ariadne* 57. <http://www.ariadne.ac.uk/issue57/stevenson/>. Archived at: <https://perma.cc/8LQ7-CQLU>.
- Suuramm, Karen. 2014. "Proposal for Implementing AtoM at UTARMS." Unpublished report.
- Trace, Ciaran B. 2020. "Maintaining Records in Context? Disrupting the Theory and Practice of Archival Classification and Arrangement." *The American Archivist* 83 (2): 322–72. <https://doi.org/10.17723/0360-9081-83.2.322>.
- Völkel, Max, Markus Krötzsch, Denny Vrandečić, Heiko Haller, and Rudi Studer. 2006. "Semantic Wikipedia." In *WWW '06: Proceedings of the 15th International Conference on World Wide Web*, 585–94. New York: Association for Computing Machinery. <https://doi.org/10.1145/1135777.1135863>.
- Wiedeman, Gregory. 2019. "The Historical Hazards of Finding Aids." *The American Archivist* 82 (2): 381–420. <https://doi.org/10.17723/aarc-82-02-20>.
- Wikidata:WikiProject University of Toronto Libraries/Discover Archives - Wikidata."n.d. Wikidata. Accessed May 6, 2022. https://www.wikidata.org/wiki/Wikidata:WikiProject_University_of_Toronto_Libraries/Discover_Archives.
- Wong, Alexandra, and Kelli Babcock. 2021. "Exploring Wikidata 'Archives At' at U of T." Presented at the LD4 Wikidata Affinity Group meeting, April 19, 2021. <http://hdl.handle.net/1807/106533>.
- Yakel, Elizabeth. 2003. "Archival Representation." *Archival Science* 3: 1–25. <https://doi.org/10.1007/BF02438926>.
- Yakel, Elizabeth, and Deborah Torres. 2003. "AI: Archival Intelligence and User Expertise." *The American Archivist* 66 (1): 51–78. <https://doi.org/10.17723/aarc.66.1.q022h85pn51n5800>.

- Yeo, Geoffrey. 2015. "Contexts, Original Orders, and Item-Level Orientation: Responding Creatively to Users' Needs and Technological Change." *Journal of Archival Organization* 12 (3–4): 170–85. <https://doi.org/10.1080/15332748.2015.1048626>.
- Zhang, Jane. 2012. "Archival Representation in the Digital Age." *Journal of Archival Organization* 10 (1): 45–68. <https://doi.org/10.1080/15332748.2012.677671>.

How to cite this article: Ansovini, Daniela, Kelli Babcock, Tanis Franco, Jiyun Alex Jung, Karen Suurtamm, and Alexandra Wong. 2022. Knowledge Lost, Knowledge Gained: The Implications of Migrating to Online Archival Descriptive Systems. *KULA: Knowledge Creation, Dissemination, and Preservation Studies* 6(3): <https://doi.org/10.18357/kula.234>

Submitted: 30 June 2021 **Accepted:** 7 February 2022 **Published:** 27 July 2022

Copyright: © 2022 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

KULA: Knowledge Creation, Dissemination, and Preservation Studies is a peer-reviewed open access journal published by University of Victoria Libraries.

OPEN ACCESS 