Toward a collaborative online framework for archival representation

by

Elizabeth-Anne Johnson

A Thesis Submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements of the degree of

MASTER OF ARTS

Joint Master’s Program

Department of History (Archival Studies)

University of Manitoba / University of Winnipeg

Winnipeg, Manitoba

Copyright © 2017 by Elizabeth-Anne Johnson
Table of Contents

Abstract ........................................................................................................................................... ii
Acknowledgements .................................. ........................................................................ iii
Introduction .......................................................... ........................................................................ 1

Chapter One: Theorizing Digitization, Materiality, and the Archives .......................................... 5

Chapter Two: Digitizing the Analogue Archives ........................................................................ 39

Chapter Three: Crowdsourcing Archival Description .................................................. 67

Conclusion ........................................................................................................................................ 96

Bibliography .......................................................... ........................................................................ 98
Abstract

Digitizing archival material is by now a standard part of archival practice. However, accurately describing and representing this material in textual descriptions is a challenge that cannot be addressed using the Rules for Archival Description, the current Canadian archival description standard. This thesis makes the case for collaborating with textual studies and digital humanities scholars to improve the framework within which descriptions are written, as well as for reaching out to members of the public and improving archival representation through crowdsourcing. The thesis also includes a case study examining the ways medieval manuscript fragments could be best represented in online description.
Acknowledgements

I would like to thank Professors Greg Bak and Tom Nesmith for their guidance and support as I make my way into the archival profession. Without their engaged and engaging discussions, this thesis would be a shadow of itself. Dr. Bak has been as great an advisor as I could ask for, and I am grateful to have had the opportunity to work with him on this project. Professor David Watt has also been influential in shaping this thesis and my own thoughts on textual studies and digital humanities.

Thanks also go to my friends and colleagues who helped keep me sane throughout the thesis-writing process, and whose insightful comments and beer-fuelled discussions of archives and academia inspire and motivate me. Olivia Weigeldt, Jocelyn Sakal Froese, Jordan Bass, Dan Elves, and Danna Slessor-Cobb have been especially kind and helpful. My family, too, has been understanding, encouraging, and supportive throughout my academic career, for which I am forever grateful. And thank you to Alex, for helping me through the final push. Jae-Sung Chon and everyone at MAKE have created the best place in the city to write, bar none, and the vast majority of this thesis was written there. Thank you.

The University of Manitoba Archives & Special Collections was my place of employment for many years and is where the seeds of this thesis were planted. Thanks to all who are and were there, particularly Brett Lougheed, for introducing me to the realities of the digital preservation of archival material.

Finally, I would like to thank the University of Manitoba for financial assistance in the form of the University of Manitoba Graduate Fellowship during the second year of my degree.
Introduction

For archives and archivists today, the push to digitize and share archival material online is growing ever stronger. The accessibility of online archival material means that it is often the material that is most used by patrons of archives. While the process of digitizing material is, by now, somewhat standardized, the long-term preservation of this material, including its description, is not. Archivists must find a way to include information about their digitized material, including its materiality and all the information contained therein, within online archival descriptions. The *Rules for Archival Description*, Canada’s national archival description standard, does not come near to serving born- or made-digital material properly, and archivists are currently working to revise it.

However, archivists as a group are overlooking a potential treasure-trove of resources. Several fields have been working with archival material in different ways: the history community’s recent work on the “archival turn” represents a new kind of critical engagement with archives. Though some of this work erases the labour and agency of archivists, it is still encouraging to see scholars from other disciplines engaging with archives in new and thoughtful ways. As historian Shannon McSheffrey points out, “thinking about how…documents were recorded and archived…destabilizes their meanings”¹. Opening archival descriptions up to non-archivists does the same thing by demonstrating that the process of describing archival material is a subjective one and that, by extension, all work done by archivists is subjective. Adele Perry, writing about the case of Delgamuukw v. British Columbia, notes that “archives are not only about what they contain within their walls. They are also about absence, although the absences in

the colonial archive are not neutral, voluntary, or strictly literal.” Making space in archival representation to discuss these absences—or at least to document the archivist’s rationale behind what material is kept and what is discarded—both disrupts the view of archival material as being the sole true record of events and requires accountability on the part of the archivist.

In addition to the work of historians, scholars and researchers in the fields of textual studies and digital humanities have been grappling with many of the archival issues of accurately portraying textual material in other formats, whether it be print or digital, for decades. By reaching out to these communities to see how they deal with the same problems, archivists can avoid having to reinvent the wheel in describing this born-digital material.

Collaborating with digital humanities and textual scholars will allow archivists to improve the descriptions they create, which will, in turn, more accurately and transparently portray the material in their collections. Material that is properly described is more discoverable, and, more importantly, it is more valuable to society as a whole. Archival description must include information about the societal provenance of the item, as well as actions taken in the process of archiving the item itself, but, when dealing with the made-digital, the materiality of the analogue item must also be documented. Scholars of textual studies have, over the years, created and refined methods of describing material items for readers who cannot work with the original item: the discipline of descriptive bibliography exists “to describe accurately the object produced by the process [of bringing a book into being] and all the variations caused by alterations in the process.”

---


and Craig S. Abbott point out, “the material forms in which texts have appeared.”

Manuscript books only exist in one singular form: there can be and often are substantial differences between copies of the same text. In light of this, and of the logistical difficulties of accessing physical copies of medieval books, textual scholars have developed ways to describe the minutest details of these books so that they can be understood by readers and scholars who are not able to study the books in person. When describing their material to an audience that is, similarly, not able to interact with it in person, archivists would be wise to take a page out of these scholars’ books.

People who work in the field of digital humanities have also spent a great deal of time and effort to develop ways to convey texts, and information about texts, online. They, like digital archivists, grapple with the issues of making digital material more accessible and of accumulating institutional support for digital projects. More importantly, however, digital humanists work on ways to preserve digital material for years to come. By looking at the preservation strategies created and maintained by the digital humanities community, archivists can learn how to look ahead, beyond the immediate action of digitization projects, and plan for the future of the digital files they create as part of these projects.

The digital humanities field is also influential in a number of crowdsourcing projects. Crowdsourcing allows archivists to include information from outside their own sphere of knowledge in their descriptions of their material—for example, Library and Archives Canada’s *Project Naming* was successful in working with Inuit people to identify Inuit individuals in historical photographs, people whose names and histories had not previously been known. Crowdsourcing has some issues—compensation for labour and the fear of waning archival authority among them—but it can also be immensely helpful when describing archival

---

4 Ibid.
collections, particularly at the item level. Crowdsourcing also engages users of archives by both making archival work more transparent and by inviting them to participate in the process of describing and archiving material.

This thesis argues for increased collaboration between archivists and the two groups of scholars mentioned above to improve the quality of online description of archival material, particularly of material which is made-digital, or which existed in an analogue form prior to its digitization. The problems of archivists are not ours alone, and it would serve our institutions, users, and materials well if we collaborated to solve those problems. As an example of this kind of collaboration, this thesis also discusses a hypothetical crowdsourcing project surrounding a collection of medieval and early modern manuscript fragments that are held in various collections at the University of Manitoba. By reaching out to communities of scholars that work parallel to archivists, we can improve the quality of our descriptions and ensure that our made-digital material is properly described and taken care of for years to come.
Chapter One: Theorizing Digitization, Materiality, and the Archives

Over the past several decades, the term “digital humanities” has come to mean much more than the computer-assisted analysis of literary texts, one of its original definitions. As more and more cultural material is born and made digital, scholars, archivists, and users of this material grapple with the meanings, uses, and consequences of the shift from analogue to digital material. As custodians and creators of some of this cultural material, archivists must devise ways to describe and preserve it. Luckily, we are not alone in this: the discrete but related fields of textual studies and digital humanities contend with many of the same issues as digital archivists do.

This chapter will compare the issues facing textual studies and digital humanities with those facing digital archival studies. Its first section will be an explanation of archival description: what its purpose is, what standards exist, how it has evolved, and how it has been and is being applied to born- and made-digital archival material. It will also explain how archival description gives information about material that bibliographic description cannot. The second section of this chapter will discuss theories of textual studies: the concept of textual material as an object, the things one can learn from the materiality of archival objects, and the way these aspects have been catalogued and described in print facsimile editions of manuscripts. The third section of this chapter will discuss the field of digital humanities and how its theories—theories of creating, interpreting, and presenting digital cultural material—can be used to bring the other two disciplines together. Digital humanities scholars investigate how creative and cultural material is digitally produced, but the interpretation of this material is also within its purview. Thinking about how best to convey descriptive information about archival material online is well within the range of digital humanities. By bringing these three fields into conversation with each
other, this chapter will provide a theoretical basis for the discussion of digital archival description that follows.

A. Archival Representation

Arranging, describing, and processing archival material is some of the most important work archivists do. Without someone taking the material donated to an archives and making it available and accessible to users, that material would likely never be seen again. The glut of information that is available to researchers today is impossible to navigate without assistance; archival description is an integral part of users’ ability to wade through the waves of information and find what they need. When writing about this process, Elizabeth Yakel chooses to use the term “archival representation” instead of the more usual “arrangement and description” or “processing,” pointing out that

[t]he term ‘archival representation’ more precisely captures the actual work of the archivist in (re)ordering, interpreting, creating surrogates, and designing architectures for representational systems that contain those surrogates to stand in for or represent actual archival materials.5

Following Yakel’s lead, and in the effort to consciously be aware of the actions and agency of the archivist, this thesis will also use the term “archival representation” where appropriate.

The need to create understandable and helpful archival representation has led to the development of several standards of description around the world. In Canada, the standard is RAD, or the Rules for Archival Description. The Canadian Council of Archives developed the Rules for Archival Description in the late 1980s and the first edition of RAD came out in 1990. RAD, which was most recently updated in 2008, was developed as an archival response to the library cataloguing systems in use in North America and Europe. These systems—the second

Edition of the Anglo-American Cataloguing Rules (AACR2) and the ISBD(G): General International Standard Bibliographic Description—had a major influence on the design and style of RAD. In the words of archivist Richard Dancy, RAD derived “most of its areas and elements of description; a certain style of writing and presentation, numbering, and punctuation conventions; division into separate media chapters; and the idea of access points and the interest in rules for the headings (names) to be used as access points” from these bibliographic description standards. While these standards were more than acceptable at the time (and for a time—AACR2 has now been replaced with Functional Requirements for Bibliographic Records, or FRBR) for the description of library material—published material that adheres to specific media designations—they are less appropriate when used to describe archival material, which by its very nature is unpublished, unorganized, and can consist of more than one media type. Describing archival material at the fonds and/or series level may require discussing material in different media, many of which may not have been published. The fields in AACR2 and other bibliographic descriptive standards are not sufficient when it comes to describing archival material.

Many aspects of archival representation do not fit into the bibliographic AACR2 paradigm. In RAD these aspects get lumped together into the Notes field at the end of a description. The lack of structure of the Notes field has been a problematic part of RAD since its inception—everything from dates of creation to notes on arrangement to restrictions on access is relegated to one massive descriptive field. Not only does this arrangement make the information in this field—information that is key to understanding the material it describes—harder to find, it

---

6 Richard Dancy, “RAD Past, Present and Future,” Archivaria 74 (Fall 2012): 10. In this thesis, the phrase “bibliographic description” will be used as above to discuss library cataloguing. When discussing the description done by bibliographers, I will use the term “textual description”.

privileges published-material-specific fields like “Publisher’s series area” over crucial archival
descriptive information. The lack of structure in RAD’s Notes field also means that it is not
machine-readable: computers are unable to sort the information in the Notes field by any kind of
semantic category, for this field is simply too broad. Archival description software like
Artefactual’s AtoM breaks the Notes field into smaller subcategories, but without such
interventions it is difficult, if not impossible, to map information from a RAD-compliant
description to another format.

It is important to include this descriptive information in archival descriptions, because it
gives researchers information about the history of the archival material up to and including the
present moment. The provenance of the material does not end when it enters an archives; rather,
every step taken by archivists, archival staff, and other users adds to the material’s provenance.
Tom Nesmith dubs the societal influences and conditions shaping the provenance or history of
the record “societal provenance,” stating that it is “not just another layer of provenance
information to add to other ones such as the title of the creator(s), functions, and organizational
links and structures. The societal dimension infuses all the others.”7 This nuanced view of the
forces that bring any archival material into being gives both archivists and users of archival
material a wider and more accurate idea of the history and meaning of the material. However, in
its current iteration, RAD does not allow for this information to be included in the body of the
description. This is particularly problematic when it comes to section 1.8B5 of RAD:
“Statement(s) of responsibility”. This section of the description makes note of “any statement(s)
of responsibility that appear outside the chief source of information or that appear on the chief

7 Tom Nesmith, “The concept of societal provenance and records of nineteenth-century Aboriginal-European
relations in Western Canada: implications for archival theory and practice,” Archival Science 6, no. 3 (December
2006), 352.
source, but not in conjunction with a formal title proper.” In other words, this is the place to identify these societal forces and anyone other than the formal creator of the records who had a hand in their creation. The sample roles *RAD* gives in this section are compiler, preparer, and photographer; other roles that can be played in the production of archival material are digitizer, arranger, and, indeed, describer. In her article “Picking Our Text,” archival educator Heather MacNeil draws parallels between the role of the archivist and the role of textual critics: “The traditional textual critic’s efforts to restore a text as closely as possible to its original, authentic form mirror the archivist’s efforts to identify and represent the original order of a body of records through arrangement.” Both textual critics and archivists are charged with shaping raw, (often) textual material into something that can easily be interpreted by members of the public; a major difference between the two fields is that textual critics seem more aware of the subjectivity of their actions. MacNeil goes on to point out that, after the archival interventions of arrangement and description, “the fonds as a whole no longer exists and what remains are the fragments that have survived, either by accident or design, out of which the archivist attempts to construct some semblance of a whole.” Framing archival arrangement and description as the (re)construction of a *fonds* makes it evident that the work an archivist does in preparing a collection for public use is complicated and should be undertaken thoughtfully.

Despite the antiseptic guidelines of *RAD*, the very act of describing archival material changes the material and the way it is perceived. The lack of defined provision for this important

---


10 Ibid., 270.
information is one of the biggest problems with RAD as a framework for archival representation. Michelle Light and Tom Hyry argue for the inclusion of colophons and annotations in finding aids, pointing out that none of the description standards they look at (RAD, EAD, and ISAD(G)) make adequate space to discuss the actions of archivists. Colophons are most often seen in books, where they identify the printer, publisher, and other information about the production of the book. Light and Hyry suggest that

archivists could use a colophon to record what they know about the history and provenance of a collection and to reveal appraisal, arrangement, description, preservation, and other decisions they made while working on a collection.\textsuperscript{11}

Adapting a finding aid—which can be read, after all, as an authored work about the archival collection—to include such a colophon would draw readers’ attention to the mediating role of the archivist. “As they selectively interpret their experiences of it,” Tom Nesmith points out, “archivists help fashion formative contexts for their work, which influence their understanding of recorded communication and position particular archives to do particular things.”\textsuperscript{12} Having to write a colophon would also make archivists more aware of the unconscious choices they all make when working to make archival material accessible.

One of the major ways archivists make their material accessible in the twenty-first century is to digitize it; this, however, results in the creation of digital files that must also be preserved. RAD is particularly ill-suited to dealing with digital media. Its ninth chapter discusses Records in Electronic Format, but does so in a way that fails to grasp the realities of cataloguing and preserving electronic records. This is particularly evident in the instructions on describing


the extent of the records (e.g. how many of them there are). The only information *RAD* requires in this instance is a description of the number of “units” (e.g. 50 emails) which, while of course crucial information when it comes to understanding the extent of an archival collection, does not come close to accurately depicting the reality of digital material. Metadata for digital material has different requirements than metadata for the analogue does: administrative metadata, which includes information about rights, preservation measures, and the creation, size, and file type of the object, is required to preserve the object, while descriptive metadata like the title of the object and the names of any people appearing within it is required to search for the object and make it accessible. As opposed to analogue records, digital records often exist in several iterations over multiple physical media and in multiple formats. As Jinfang Niu points out, “[w]hen records that belong together conceptually are separated physically, they need to be conceptually integrated and described together in archival finding aids, so that users can see the conceptual relationships among the records.”\(^1\) Simply stating that a collection contains a webpage, for example, obscures the fact that said webpage is comprised of many objects in many formats—code, style sheets, text, and photographs, to name a few. Similarly, an email may contain a single word of text or a series of high-resolution photographs. Describing the extent of digital material in the way *RAD* mandates does not give archivists or researchers any real information about the size of the collection in terabytes or about the requirements to preserve or view the material. Sections 9.5B2 and 9.5B3 discuss optional approaches to describing electronic records, suggesting that archivists give the “digital extent” (e.g. 1.5 GB) of the archival material and the number of physical carriers (e.g. 2 CD-ROMs). However, if archivists do not describe all three of these properties, researchers will not be able to accurately gauge what exactly is included in the archival

collection. Furthermore, RAD does not require its users to include the file format of an electronic file in a description. While its requirement that a filename not be used as the official title of an item is understandable, RAD’s laissez-faire approach to recording filenames and extensions at all is troubling. Particularly because preserved digital material may be viewed in a different format from that in which it is preserved (e.g. a .gif file that is ingested into a digital preservation system may be preserved as a .tif file and then viewed as a .jpg), it is important to document not just the preservation format of the digital material, but also the other formats into which it may be or may have already been migrated. Without preserving file type information, the process of preserving digital material, which often involves the migration of files from one type to another, is needlessly complicated. Having the ability to see at a glance which file types a fonds contains would streamline the process of selecting fonds for preservation and, if the patron has an idea of the contents of the collection, the process of selecting fonds for access.

Effectively preserving digital archival material also requires more information than simply file types. PREMIS (Preservation Metadata: Implementation Strategies, a working group and data dictionary that defines metadata elements required for digital preservation) explains that “preservation metadata supports activities intended to ensure the long-term usability of a digital resource.” These activities can include using checksums to make sure a file has not changed over time, migrating a file forward to a format that is approved for preservation, and recording a file’s provenance through any changes in format or custody.


In order to determine which fonds to examine, archival users must know more about what those fonds contain than *RAD* currently requires. Users need to know which file types a fonds contains, and the file types from which those items may have been migrated. They need to know the technical details of any migrations in format, and whether and why any such migrations have taken place at all. They need to be assured through the use of checksums that the files listed in the finding aid are, in fact, the ones that exist in the collection. By not requiring this information to be included in descriptions, *RAD* is doing many of its users—and researchers across Canada—a disservice.

The inclusion of information about the history of the archival item within the archives is also crucial to understanding the item. This includes information about whether or not—and how, and when—an archival object has been digitized. For specialist users, like academics and other archivists, it is important to know the contexts of the digitization process, as well as the history of the material prior to its digitization. The camera or scanner used, and any proprietary formats created by the capture hardware; the lighting setup; and the processing of digital files from the camera to the archives’s website or server all have an effect on the way the material is seen and interpreted by researchers. If all this information were organized in a separate field in the description, rather than being relegated to the disorganized Notes field (if it is recorded at all), researchers and computer systems could easily find and use it. The majority of archival material does not have a publisher, let alone a place of publication, so why are there dedicated spaces in *RAD* for this information? Information that is actually important to the contexts and understanding of archival material must be given a dedicated section of archival description.

---

16 Digital cameras often use a proprietary “camera raw” or digital negative format that contains the raw data captured by the camera. These files can be normalized to the open DNG format for preservation, but this normalization must be documented.
Describing the place where archival material was published is detailed in section 9.4C of *RAD*. The availability of other formats—like the digitized version of archival material—is buried in section 9.8B16b. This is just one example of the unbalanced priorities of *RAD*, particularly when it comes to born- and made-digital material. Researchers, archivists, and the material itself all deserve better.

In early 2016, the Canadian Council on Archival Description held a national meeting on the future of *RAD*. Taking into account responses to a nationwide survey of archivists about *RAD* and its efficacy as a description standard, CCAD made a number of recommendations, some of which are quoted below:

1. Retain a Canadian national archival descriptive standard (RAD3)
2. Commit to the general principle of keeping RAD aligned with the ICA’s international archival descriptive standards; any departures from those standards…must be explicitly noted and justified.
12. Study the requirements for description of electronic records with the goal of identifying a core set of descriptive elements for digital archival materials.
18. Study options for opening Canadian descriptive systems to participatory archives and how this might affect RAD.\(^\text{17}\)

These recommendations are promising. If they are followed, unlike the recommendations made in the attempt to create *RAD2* in 2005 which largely were not, the Canadian national descriptive standard may finally be in line with (and in some respects better than) international standards.

For example, although another description standard, *ISAD(G)*, is an improvement over the current iteration of *RAD* in that it features more discrete fields for information, it too disregards the importance of post-acquisition provenance. *ISAD(G)* does have an element called “Archival history”, the purpose of which is “[t]o provide information on the history of the unit of

---

description that is significant for its authenticity, integrity and interpretation.”18 This field is supposed to include “those actions, such as history of the arrangement, production of contemporary finding aids, re-use of the records for other purposes or software migrations, that have contributed to its present structure or arrangement.”19 However, as Heather MacNeil has noted, if the “unit of description is acquired directly from the creator,” this field can be skipped entirely in lieu of the “Immediate source of acquisition” field.20 This field “makes no provision for recording the history of the records’ arrangement and representation,” allowing the significant changes made to the material in the process of arrangement and description—and the active work of the archivist—to disappear.21

A crucial part of the active work archivists do is digitizing archival material to preserve it and make it more accessible. To discuss this digitized archival material properly, it is also necessary to discuss the process of digitization. This act, which can easily be taken for granted, has profound effects on the archival object and the way researchers interact with it. By necessarily transforming the physicality of interacting with the artifact and replacing it with a physical experience that is mediated by a screen, digitization creates an entirely new archival object. Other archival processes, such as format migration (converting archival material from one format to another, whether from one digital format to another or from printed text to microfilm) also create new archival objects; all of these actions must be documented. These transformations...


19 Ibid.

20 Ibid.

are not inherently good or bad things, but they are the results of choices the archivist make. Users should be able to see clearly which choices archivists made in the preservation of archival material and understand how these changes affect their access to records.

By far the biggest advantage of digitization is improved access to material. With rare books and manuscripts available online, anyone with Internet access can read and interact with texts to which they would otherwise have to travel. Many of these texts are also held in institutions that require credentials to get into; the online dissemination of digital facsimiles of rare books and manuscripts serves to democratize the information held in these books. Indeed, as Ian Milligan points out, many analogue archival resources are now less accessible than they once were. In 2012, Library and Archives Canada (LAC) entirely eliminated walk-in reference services; researchers must now make an appointment in advance if they wish to meet with a member of the reference staff. This change was made as part of a shift to an online service model where, according to LAC, “all Canadians will be able to discover, engage with, and share LAC’s rich content when, where, and how [Canadians] want it.” As of April 2017, the reference rooms at LAC now provide service for five hours each weekday; rather than being welcomed in at any time during LAC’s hours of operations, patrons are still encouraged to make appointments with archives staff in advance or to ask questions online.

---


24. Ibid.

wanted to engage with archival material solely through the Internet, however, most users of archives will have no other choice. Cuts in funding to heritage institutions like LAC mean that researchers may not be able to meet with archival staff even if they do have the means to travel to an archives; cuts to the educational institutions that fund much research work mean that researchers simply do not have the means to travel to archives in the first place. Milligan also points out that pressure to increase time-to-completion rates for graduate students “may also be encouraging students to move toward greater numbers of online sources. They are certainly more cost effective than distant traditional sources.”

Putting digitized versions of archival material online makes the information in it available to more people and, more importantly, to more people who are not necessarily academic specialists. It also provides a way for members of the public to view and interact with archival material that may be too fragile to circulate normally. For example, the 2015 launch of the fourth edition of *Electronic Beowulf*, a free online facsimile and edition of MS Cotton Vitellius A. xv, the only extant (and much-damaged) manuscript copy of *Beowulf*, allows scholars and other interested people from all over the world to work with a manuscript that is too fragile to be worked with in person. Earlier editions of *Electronic Beowulf* were only available for purchase on CD-ROM or DVD; however, security problems with the programming of the physical media editions of the project led editor Kevin Kiernan and programmer Emil Iacob to publish the fourth edition online in October 2015. Through digitizing, editing, and uploading the manuscript, the

---


Electronic Beowulf project has allowed people from all over the world to access the manuscript of this seminal English text while preserving the original manuscript from further damage.

Another major advantage of digitization is the ability to work with images of artifacts to discern information about them that is hidden when working with physical artifacts. As Carolyn Steedman noted in her article “After the Archive,” digital images can be manipulated in ways that the analogue book cannot.\(^{29}\) Indeed, with the advent of smartphones, nearly every researcher who does have the resources to travel to archives in person has a camera in their pocket. When working with digital images of texts, researchers can zoom in on, change the contrast of, and examine books and pages in new ways. If a word or letter in a manuscript is unclear, a high-resolution image of the page may reveal things that the naked eye can miss. Individual pages can be compared to each other digitally in ways they cannot be in the analogue world; for example, looking at pages from the same book side-by-side and comparing similar passages from different versions of the text that are held separately from each other are both possible with digital facsimiles. The Archimedes Palimpsest is an example of a book that has been made substantially more meaningful by digital technology. A palimpsest is a document from which the original text has been erased so that the writing support material can be reused. This palimpsest is a thirteenth-century prayer book that contains, underneath the later religious writings, erased copies of tenth-century texts by the Greek mathematician Archimedes. By using multi-spectral imaging as well as other ways of looking at the images of the pages of this text, scholars with the Walters Art Museum were able to regain access to Archimedes’ words.\(^{30}\)


made the recovery of these texts possible. The *Electronic Beowulf* edition mentioned above also uses digital technologies to recover fragmentary text. One leaf of the *Beowulf* manuscript is also a palimpsest; its obscured text is rendered even more confusing by the possibility of ink from the facing page (page 178v) being off-set onto the palimpsest page (179r). The digital edition of the manuscript allows the user to overlay a reversed image of 178v onto the image of 179r, providing a way to match up the remaining marks on 179r with the possible off-set letters from 178v.\(^{31}\) Tools like this one allow users of digitized texts to attempt to restore text that was previously destroyed.

When working specifically with fragments, it is even possible, to an extent, to reassemble books that have been destroyed. Otto Ege was a scholar and self-defined “biblioclast”, or book-destroyer.\(^{32}\) During the early twentieth century, Ege systematically destroyed dozens of medieval manuscripts in order to divide them into individual leaves, which he then collected and sold to various collectors and institutions. Ege proudly identified himself as a biblioclast, an “‘aesthetic ghoul’ of the book world,” stating

…Surely to allow a thousand people ‘to have and to hold’ an original manuscript leaf, and to get the thrill and understanding that comes only from actual and frequent contact with these art heritages, is justification enough for the scattering of fragments. Few, indeed, can hope to own a complete manuscript book; hundreds, however, may own a leaf.\(^{33}\)

These collections, titled “Fifty Original Leaves of Medieval Manuscripts,” can be found across North America and around the world. One box is held at the University of Saskatchewan, where

---


\(^{33}\) Ibid., 517-18.
Professor Peter Stoicheff is working on rejoining his institution’s pages with their siblings.

Stoicheff discusses the digital and virtual reconstruction of Ege books in “Putting Humpty Together Again: Otto Ege’s Scattered Leaves”:

Much as we would love to have people send us their actual Beauvais Missal [known as one of the most beautiful books in the Ege collection] leaves, the fact that the collector in Chicago with Box #1 won’t even reveal his name suggests it’s not likely. But the digitized edition of the Beauvais Missal would have its benefits beyond the physical book. Its availability on the web would complement rather nicely Ege’s original motive for dismembering books and selling the boxes. Beauvais Missal leaf owners, many of whom are unaware of the original book context of their possession, could have the opportunity to collaborate in the project of recreating it.34

Stoicheff is aware of the limitations of reconstructing these books physically—in addition to their value and their owners’ reluctance to give them up, there are pieces of the manuscripts, such as their bindings and colophons, that are lost and may never be recovered—and of the value of bringing these orphaned leaves back together digitally. The reconstructed books will not be the same as they were before Ege cut them apart; how could they be? But by making digital analogues of these leaves, people will be creating a new way of looking at them and a new edition of the books that had been destroyed. Bibliographic description gives us a window into the past lives of these texts.

As bibliographic description informs how we read editions of manuscripts, archival description is the lens through which all users of archives initially view archival material. The influence that a description has on users’ interactions with archival material must not be understated; on the contrary, archivists must provide as much contextual and descriptive information about archival material as possible so that researchers can find the material they are

seeking. The *Rules for Archival Description* allows archivists to provide some of this information, but it does not nearly meet the needs of archival material. Although the Canadian archival community has, in the past, been slow to come to a consensus on changing the national standard, the most recent survey on the topic conducted by the Canadian Committee on Archival Description had a much higher response rate than earlier discussions of the same topic. This standard, if it is to adequately meet the needs of archivists and other users of archives, must include discrete spaces for information about the provenance, transmission, and archival custody of the material, including its handling at the archives; information on the current format of the archival material and on any other formats it has been in; and information about the archivist and/or staff person who prepared the description. These elements must also be easily mappable to other standards; being able to crosswalk the metadata stored in an archival description is crucial to making that data accessible to people using different description standards. By clarifying the history of the record, and by identifying the forces that made the record what it is today, archivists can provide users of their archives with a much more comprehensive understanding of their records.

35 “However, the ambitious restructuring proposed in 2004 as RAD2 was greeted by Canadian archivists with little consensus and some hostility against a general backdrop of indifference; it was not implemented.” Richard Dancy, “RAD Past, Present, and Future,” *Archivaria* 74 (Fall 2012): 9.

B: Textual Studies

Documenting the history of a record is something that is familiar to textual scholars. Their field, called textual or bibliographical studies, is an old one, dating (at least) back to the Library at Alexandria, but still relevant today. Archivists working with representation particularly can learn from textual scholars’ experience conveying the physicality of their texts by describing them. Textual studies, as is perhaps evident from its name, looks at texts and the physical ways they are created, disseminated, and read. The physical nature of texts, the marks on the paper, the paper (or parchment, or papyrus) itself are the foundation for every other kind of interpretation of the texts. Indeed, in his “History of textual scholarship,” David Greetham goes one step further and explicitly compares textual transmissions to other, older transmissions: “Anyone who has played the game of ‘Telephone’…will know that variance is an inevitable part of any transmission—for good or ill.” In this section I will discuss the ways textual scholarship deals with these inevitable variances, and how archival studies might learn from them.

When introducing the concept of textual studies, William Proctor Williams and Craig S. Abbott begin with the simple sentence: “Texts have lives.” The lives texts live, the way they are created, and how they come to be read are all concerns of textual scholars. Parallels may be drawn between this method of looking at texts and Nesmith’s theory of societal provenance as mentioned above. There is much more to reading than just looking at words on pages, just like


there is much more to archival material than just a pile of papers in a box. In order to go beyond
the surface, it is necessary to look at the history and provenance of the texts.

Historical bibliography is concerned with more than the provenance of a text; this field
looks at the creation of a text and its journey from its author’s mind through drafts, proofs, and a
“final” published version. It can also focus on specific aspects of book production, such as
bindings, papers, or handwriting (the study of which is known as palaeography). In the absence
of front matter or colophons, as is often the case with both medieval manuscripts and archival
material, having knowledge of these aspects of text production can provide useful clues to the
provenance and history of the text. In both these cases, the front matter may be missing because
it was never included with the text—the book may lack them, or the record may not record the
information—or because it has, over time, become separated from the main text and lost.

Understanding the materiality of these texts is a crucial part of understanding them and
their place in the world. According to Williams and Abbott, this form of bibliography comprises
both analytical and descriptive bibliography, which can be lumped together under physical
bibliography. Analytical bibliography is interested in the processes of the production of the
book, while descriptive bibliography focusses on the accurate description of those processes.
Bibliography is a sister discipline to diplomacy: as Luciana Duranti points out, “diplomatics
studies the written document, that is, evidence which is produced on a medium (paper, magnetic
tape, disc, plate, etc.) by means of a writing instrument (pen, pencil, typing machine, printer,
etc.) or of an apparatus for fixing data, images, and/or voices.” Duranti goes on to state that

40 Williams and Abbott, Introduction to Bibliographical & Textual Studies, 11.
original.
“[t]he form of a written document is, therefore, the whole of its characteristics which can be separated from the determination of the particular subjects, persons, or places it is about.”42 Physical bibliography is equally concerned with the physicality of documents, but unlike Duranti’s interpretation of diplomatics, it has a broader scope than simply “archival documents” “created or received by a physical or juridical person in the course of a practical activity…[and not] expressing feelings and thoughts and created by individuals in their most private capacity.”43 In other words, bibliography, like diplomatics, deals with the forms of texts, but does not limit itself to purely “archival” texts or strictly categorize texts based on their forms.44 Understanding these forms is important: creating an accurate description of what a text is and how it was created requires the ability to read the text on a deeper level to discover clues to its history, and theories of textual scholarship provide a way to gain this ability without eliminating some archival material as not “archival” enough.

Among the information that can be gleaned by examining texts bibliographically is knowledge of the text’s age, place of origin, and the social context of its creation. For example, if one looks at a text palaeographically, one can discern its general age and physical origin. Having the ability to read into the text on a different level than simply reading the words means that one can gather information that might otherwise be forgotten.

The same can be said for the ability to look at the physical supports of the text—that is, the material on which it was written and in which it was bound. Is there a stretched-out hole in

42 Ibid.
43 Ibid.
44 In this context I use “archival” to refer to its use in Jenkinsonian and neo-Jenkinsonian writing, such as Duranti’s. Hilary Jenkinson defined archives narrowly as “documents which formed part of an official transaction and were preserved for official reference.” Hilary Jenkinson, A Manual of Archive Administration (London: Percy Lund, Humphries and Co., Ltd., second edition, 1937), 4.
the parchment? The cow from which that page was made might have had a bug bite. Does the
parchment have a thick, discoloured line in it? That is where the animal’s spine once was. Being aware of these physical aspects of textual material means that a reader can have a greater
connection with the text, but also that the text is situated in reality, in a specific time and place
that are a crucial component of the work’s societal provenance.

The field of textual studies also encompasses the process of textual criticism and the
process of creating an edition of a text. In the context of textual criticism, an “edition” is a
version of a work that has been edited by scholars. From here, though, there are several types of
editions. Single-manuscript editions are based on one manuscript, as opposed to critical editions
that use more than one source. Many single-manuscript editions are also diplomatic editions,
where effort is made to recreate all of the original page’s apparatus (marginalia, punctuation, line
breaks, etc.) on the printed page of the edition. This kind of edition is the one that has most
fallen by the wayside in the age of easy photographic and digital reproduction, which have
simplified the production of facsimile editions that are intended to reproduce as closely as
possible the physical experience of paging through the original manuscript. Critical editions do
not reproduce a particular instance of a text, but rather construct “a text that may incorporate
readings from several documentary texts and may include editorial emendations that establish

45 Clemens and Graham, 11-12

46 The textual studies term “diplomatics” is related to but distinct from Luciana Duranti’s use of the same term in
archival theory as discussed above. Diplomatics has traditionally been the study of the physical form of documents,
often in order to determine the authenticity of the documents. Diplomatic editions of texts attempt to reproduce a
single manuscript text in type, including the original reading of the text, revisions made by scribal correctors, and
principles, concepts, and methods of diplomatics…can bring system and objectivity to archival research into
documentary forms.” (“Diplomatics: New Uses for an Old Science,” 8). Duranti’s approach to authenticity is
discussed further in chapter three of this thesis.
readings not found in any document.”  

Finally, hypertext editions can be any of the above, as long as they make use of the different ways hypertext allows text(s) to be presented. The work archivists do when they digitize rare books, manuscripts, and records is that of creating a facsimile edition. The problem with this is that the other important factor of an edition—the discussion of what has been done and why, the description of the manuscript’s history and provenance—is often missing.

The process of retransmitting a text by creating an edition of it is something that textual scholars take very seriously. Transferring text from one medium to another, from handwriting on parchment to print on paper, for instance, or from parchment to digital storage to web to screen, inevitably creates a new version of the text and the physical artifact that conveys it. This recreation has profound impacts on the way the text is presented, interpreted, and read. As Michelle R. Warren put it, “[m]aterial forms…structure a number of different relations of ‘power’—between editor and text, edited text and source materials, edited text and readers, different parts of the edition itself.”  

Scanning a manuscript and uploading it to an archives’s website does more to a text than just make it accessible in a new way. Those doing the scanning and uploading are forcing the users of the resource they create to mediate the text differently, to interact with it in a new way, to consciously or unconsciously deal with the text as something that has been changed in order that they might see it. They must also be responsible for describing and justifying the new text they have created.

---

47 Williams and Abbott, *Introduction to Bibliographical & Textual Studies*, 78.

Descriptive bibliography, as discussed above, is the process of creating as detailed a description of the text’s physical form as possible. Williams and Abbott discuss the seven questions that a textual description must answer:

1. What is the book—what edition, for example? ...
2. What does the book say about itself in the title page, colophon, copyright page, and other imprints?
3. How was the book put together? Considered here are the imposition format, collation formula, and pagination.
4. What does the book contain?
5. What is the book made of? What sort of type [or script] and paper [or parchment] were used?
6. How is the book packaged in a binding and dust jacket?
7. What is known, from bibliographical analysis and other sources, about the printing and publishing of the book, its variant states, and the irregularities of particular copies?

These seven points are all relevant for both users of archives and for archivists who manage archival materials. There are clear parallels between this textual description and the archival description laid out by RAD. Indeed, archival arrangement and description can go one step further than descriptive bibliography when it looks at the function of the texts and the reasons behind their creation. While there is no ruling body that defines what must and must not be in a textual description the same way there is for archival and library descriptions, there are commonalities between various scholars’ descriptions and conventions that have arisen over the years. Books like D.C. Greetham’s *Textual Scholarship: An Introduction*, G. Thomas Tanselle’s *Rationale of Textual Criticism*, and D.F. McKenzie’s *Bibliography and the Sociology of Texts* record, promote, and pass on the ways different scholars approach their texts.

These scholars approach their texts differently from archivists, but share many of the same aims. The similarities between and common genesis of textual bibliography and archival diplomatics as discussed by Luciana Duranti speak to this and to the value of interdisciplinary collaboration between archivists and textual scholars. By discussing the description of archival

---

material with others who do similar work, archivists can improve their own methods of description.

C: Digital Humanities

Although it was not developed specifically with archival concerns in mind, the field of digital humanities may have answers to many of the questions that archivists and textual scholars pose. Because digital humanities projects can deal with theories of creating, interpreting, and presenting digital cultural material, its practitioners have been grappling with the same issues presented by made-digital archival material as archivists have. Matthew Kirschenbaum explains the field of digital humanities as being “about a scholarship (and a pedagogy) that is publicly visible in ways to which we are generally unaccustomed, a scholarship and pedagogy that are bound up with infrastructure in ways that are deeper and more explicit than we are generally accustomed to.” While archivists may be less explicitly focussed on scholarship and pedagogy than the teaching academics Kirschenbaum is addressing, the public nature and infrastructure reliance of digital humanities projects will be familiar to those working with digital archives exhibits and preservation. While the specific topics of study, research, or work we do may vary, digital humanities scholars and digital archivists grapple with some of the same problems: creating access to digital material and gaining infrastructure support for the continuing project of maintaining this material over time. Digital archivists could well be described as being part of the larger digital humanities community, and it is essential that we communicate.

The digital humanities community is as broad as its name suggests: as Kathleen Fitzpatrick states in her 2012 article “The Humanities, Done Digitally,” the field of digital

---

humanities is “broadly humanities-based and includes scholars in history,…media studies, and other fields that can benefit from bringing computing technologies to bear on traditional humanities materials.”  

Indeed, Fitzpatrick explicitly mentions that some digital humanities projects have been “archival in nature” but stops short of discussing what, exactly, makes them archival or how digital humanists and archivists collaborate.

The field of digital humanities emerged out of the growing use of computing technologies in the humanities. The Association for Computers in the Humanities and the Association for Literary and Linguistic Computing began holding joint conferences in 1989; these conferences evolved into the annual Digital Humanities conference, which is still being held. In his introductory essay in Debates in the Digital Humanities, Matthew Kirschenbaum defines digital humanities as being “more akin to a common methodological outlook than an investment in any one specific set of texts or even technologies.” This way of looking at digital humanities underlines its flexibility and use as a framework rather than as a hard and fast series of critical techniques. Kirschenbaum goes on to mention a variety of projects that fall under the umbrella of digital humanities, from mass text analysis to the creation of digital facsimiles to the

---


52 Ibid. It is possible that Fitzpatrick is using “archival” in the way mentioned by Joan M. Schwartz and Terry Cook: “While some writers have begun exploring aspects of ‘the archive’ in a metaphorical or philosophical sense, this is almost always done without even a rudimentary understanding of archives as real institutions, as a real profession (the second oldest!), and as a real discipline with its own set of theories, methodologies, and practices.” (Joan M. Schwartz and Terry Cook, “Archives, Records, and Power: The Making of Modern Memory, Archival Science 2 (2002): 2.) Kate Theimer also discusses the broadening definition of “archive” in the Journal of Digital Humanities: “…in the broadening of ‘archives’ to extend to any digital collection of surrogates there is the potential for a loss of understanding and appreciation of the historical context that archives preserve in their collections, and the unique role that archives play as custodians of materials in this context.” (Kate Theimer, “Archives in Context and as Context,” Journal of Digital Humanities 1, no. 2 (Spring 2012), accessed May 10, 2017, http://journalofdigitalhumanities.org/1-2/archives-in-context-and-as-context-by-kate-theimer/.)

preservation of digital media like video games and virtual communities. Communities themselves are an integral part of digital humanities: it is a field that has been conscious of collaboration from its very early days. Bridging the two disparate fields of humanities and computing required and requires scholars to work together to create not just a system but a language to communicate each party’s needs.

As can be seen by the above examples, digital humanities has been and will continue to be defined in many different ways. Indeed, the website “What Is Digital Humanities?” displays one of 817 definitions of the field, contributed by people working in digital humanities, whenever it is refreshed. However, Lisa Spiro points out that “[r]unning throughout these statements is an overarching sense that the digital humanities should promote traditional humanistic values such as access to knowledge and civic responsibility by embracing collaboration, cross-disciplinarity, innovation, participation, and openness.” These values are integral to much current archival thinking, particularly around collaboration, participation, and openness; however, they also demonstrate the incredible breadth of work being done within the digital humanities. For the purposes of this thesis, I will focus on digital humanities projects that

---

54 Jason Heppler, “What Is Digital Humanities?,” accessed May 10, 2017, http://whatisdigitalhumanities.com/. Some sample definitions, gathered from participants in the annual “Day of DH” project, are “The development, exploration, and evaluation of computer-based technologies and resources for enabling the pursuit of research questions in the humanities,” attributed to Susan Brown; “Humanism and its universe, digitally,” attributed to Guyda Armstrong; “On one level, digital humanities is about applying computational methods to questions and problems that have traditionally interested scholars in the humanities. Computational analysis can make it possible to tease out patterns or relationships that are difficult, or impossible, for humans to detect. The purview of the digital humanities extends beyond methodology, however. It is also about raising new questions and exploring topics such as digital literacy and the cultural impact of new media. Whatever the approach, academic libraries have an important role to play in providing the infrastructure and support that digital humanities projects require in order to be sustainable,” attributed to Tim Thompson; and “As a designer attached to a department of academics I wouldn't even start,” attributed to David Little.

investigate collaboration and participation, particularly in the area of crowdsourcing, and on areas of innovation around digitization and digital preservation.

One of the ways digital humanities scholars research collaboration and participation is by studying user experience. Charlie Edwards investigates user experience in “The Digital Humanities and Its Users” and discusses the divide between skilled and unskilled users of digital humanities projects. The technical barriers to participating in some forms of digital humanities work are not insignificant, but Edwards points out that “[t]wo of DH’s most recent and most successful projects, in fact, are specifically aimed at engaging the unskilled.” Edwards is referring to the Transcribe Bentham and DHAnswers projects, which both make use of “amateur” participants. However, Edwards is likely underestimating both the abilities and skills of participants in crowdsourcing projects. As Daren Brabham points out, and as will be discussed further in chapter three of this thesis, the crowds that do crowdsourcing work tend not to be amateurs, but experts who choose to participate in these projects for a variety of reasons. Regardless of the skill level of crowdsourcing volunteers, ensuring a smooth user experience is key to attracting and maintaining participants. Transcribe Bentham is a project at University College London that “aims to harness the power of crowdsourcing to complete the transcription of 12,500 of Jeremy Bentham’s manuscripts.” By opening up the work of transcribing these manuscripts to the public, the researchers behind the Transcribe Bentham project get free labour


(a problematic aspect to crowdsourcing that will be also discussed in chapter three) while also
demolishing the “ivory tower” in which academic work too often resides. The other digital
humanities project Edwards mentions is *DHAnswers*, where users of the Association for
Computers and the Humanities website (it, along with the *Chronicle of Higher Education’s*
“ProfHacker” blog are the powers behind the site) can pose questions to people who work in the
digital humanities. As Edwards puts it, this agency to ask questions empowers those doing the
asking to help shape the field: “those asking the questions act as strong levers inducing the
community to document its knowledge.”59 Digital humanities projects can be fluid and
collaborative by design: users with any background can contribute and affect the end result. By
opening up archival description in the same way that these projects have opened up the field of
digital humanities, archivists can gain information about archival material that they may not
already know as well as the users do and demystify the process of archiving. By involving users
in a more active role, digital humanities projects have made great gains. If more archives follow
their lead, like institutions like NARA and the Library of Congress already have, they too can
increase user engagement, gather new information about their holdings, and make their programs
and mission clearer to more people.60

A major throughline in digital humanities is the continuum between creating,
interpreting, and preserving material. Preserving material allows it to be read and interpreted at a
later date; in a certain sense, texts are only created when they are read.61 Sue McKemmish
discusses this conceptual idea of texts in the context of archival description, pointing out that


60 Michelle Springer et al., *For the Common Good: The Library of Congress Flickr Pilot Project*, (Washington, DC:

61 That is, because each person experiences the archival object in a unique and personal way, the artifact itself is
different to each viewer. Postmodern archivarial theorists like Tom Nesmith gesture towards this concept: Nesmith
[e]ven when documenting records in traditional forms, archival systems cannot fulfil this purpose [ensuring that records are preserved in the context of creation and use, and that they retain their qualities as evidence] if they do not go beyond concerns with the physical grouping and description of records in the repository, to capture data about contextual and documentary relationships.  

The ways this continuum intersects with archival theory and action are interesting; in a field so concerned with the keeping and preservation of relevant, important material, what exactly are archivists doing when they select material as worthy of preservation and create a new, digital analogue of the original material? Terry Cook, writing about archival appraisal, points out that

[a]rchivists inevitably will inject their own values into [archival] activities…by their very choice, in eras of limited resources and overwhelming volumes of records, of which creators, which systems, which functions, which transactions, which descriptive and diffusion mechanisms, indeed which records, will get full, partial, or no archival attention.  

The power archivists have to shape which material survives and which does not is often elided in discussion of archival methods. However, the very act of “archiving” material transforms it, giving it a new meaning and value. “As they make determinations about archival or historical value, archivists in effect create, initiate or perpetuate an axiological commitment which is manifested in the permanence of the order that emerges,” to quote archivist Brien Brothman.  

The work archivists do in deciding which material survives and which is discarded has a

points out that “A record is a meaningful communication, which means it consists of a physical object, plus an understanding, or representation of it. Some of what makes a record meaningful is inscribed within it, but often much of what makes it intelligible is not.” (Tom Nesmith, “Still Fuzzy, But More Accurate: Some Thoughts on the 'Ghosts' of Archival Theory,” Archivaria 47 (1999), 144.)  


permanent effect on the historical record. While choices about digitization may seem less permanent, they have a large effect on which records are the most discoverable and accessible. Digital humanities theories can provide insights into this problem, which archivists must be prepared to solve. Indeed, archival studies and textual studies intersect at this point: they are both concerned with making texts available in different ways for long periods of time and must contend with the implications of this desire.

The problem of digital preservation is one with which digital humanities scholars also contend. The seminal digital artwork *Agrippa (a book of the dead)*, written by William Gibson, illustrated by Dennis Ashbaugh, and published by Kevin Begos Jr. in 1992, engages with this problem head-on: the book was published on a 3.5-inch floppy diskette together with a program that “devours the text as you read it,” and the poem was designed to destroy itself during its first reading.65 In his book *Mechanisms*, Matthew Kirschenbaum cites *Agrippa* as a reminder that “preservation is ultimately a social domain, where actions and agency can serve to trump purely technical considerations.”66 *Agrippa* does survive today; a reproduction of it can be viewed on YouTube67, and various transcriptions and emulations can be seen on *The Agrippa Files*, a website created by members of the University of California, Santa Barbara Transcriptions Project.68 Kirschenbaum’s discussion of *Agrippa* (and his contributions to *The Agrippa Files*) is


an excellent illustration of the ways digital humanities scholars concern themselves with digital preservation. *Agrippa* was never supposed to survive, but because of its impermanent design (and, as Kirschenbaum points out, a dearth of actual physical copies of the book) and the challenge it posed, the poem found renewed life on the MindVox BBS and, thence, to other early Web sites.\(^{69}\) Although the example of Kirschenbaum and *Agrippa* is less an example of digital humanities scholars doing preservation than studying it, the work being done on digital material by non-archivists can and does have an impact on the preservation strategies implemented by archivists. Understanding how digital cultural material is used, transmitted, and transformed is key to preserving it so that it can be accessed by future generations. As Kirschenbaum concludes in his introduction,

> new media cannot be studied apart from individual instances of inscription, object, and code as they propagate on, across, and through specific storage devices, operating systems, software environments, and network protocols; yet the forensic imagination of the book’s subtitle [*New Media and the Forensic Imagination*] is also conceived as a deeply humanistic way of knowing, one that assigns value to time, history, and social or material circumstance—even trauma and wear—as part of our thinking about new media.\(^{70}\)

Archivists have become well-versed in recognizing the values of time, history, and circumstance as they affect textual and other analogue records. We must also take these factors into consideration when working with digital material if we want to preserve it in its proper context.

**Case Study**

Throughout this thesis, I will be applying my discussion to a concrete project: the medieval manuscript fragments at the University of Manitoba. These fragments are currently part of a project that Dr. David Watt and I are working on. They are a collection of thirteen

---


\(^{70}\) Ibid., 23.
fragmentary texts dating from the tenth to fifteenth centuries. Rather than existing on their own, most of these fragments are part of the bindings of early printed books; once these books made handwritten parchment books obsolete, the manuscript books were often reused as sturdy binding supports for early printed books. Hence, fragments of manuscripts can often be found in the bindings of these books, and can give clues as to the production and provenance of incunables. These are small fragments, only a few centimetres in length and width. In addition to their small size, the content of the fragments may be obscured with dirt or other binding apparatus. This makes identifying and cataloguing them challenging. These small fragments are held in the University of Manitoba Archives & Special Collections: they and their descriptions straddle the line between archival and textual/bibliographic work.

On the other end of the spectrum are the fragments held in St. John’s College at the University of Manitoba. Rather than being used for their material strength as the binding fragments were, these fragments were removed from their books and framed, used as decoration. They too lack context, but are in better condition and easier to read than the binding fragments. The St. John’s fragments also have identifying plaques hanging with them, which help trace their provenance. Because none of these fragments were catalogued before our project, they are excellent candidates for cataloguing and description from a blank slate.

Much like more traditional archival material, these fragments do not have a known publisher or place of publication; this underscores the similar demands of archival and bibliographic description. Particularly with the binding fragments, the only clues we have to their history are the book they are bound in and whatever clues we can glean from the handwriting itself. One of the manuscript fragments held at the University of Manitoba Archives & Special Collections can be dated with relative certainty to the ninth century, based on the shapes and
sizes of the letters. The same clues lead one to believe it was created in Italy—information which is backed up by the archives’s records of the manuscript in which the fragment was found. As discussed above, information on the hand(s) in which material is written is among the data not easily incorporated into the RAD descriptions used by archives. Digitizing these fragments would allow them to be ingested into a system that allows members of the public to contribute information that Dr. Watt and I may have missed to the fragments’ description, improving our understanding of the fragments and their history. Improving the descriptive process in use at the University of Manitoba will also allow this to happen; as discussed in chapter three, crowdsourcing is a way to reach beyond the limited knowledge of archivists and other information professionals and gather information from a wider range of people.

Conclusion

Archivists are rightly concerned about the future of digital archival material: this material may share some content with the traditional analogue media forms archivists have dealt with, but it requires a much different level of care. One of the ways archivists care for archival material is describing it so that it can be accessed by users; the Canadian archival description standard, *Rules for Archival Description*, ensures that archival material across Canada is described in similar ways that can be understood by users of any archives in the country. However, digital material is currently underserved by *RAD*, and archivists must come up with a better way to make their digital holdings discoverable and useable. Another factor that archivists must take into consideration is the transmission of the important physical aspects of made-digital representations of analogue materials. Luckily, archivists are not alone in this endeavour. The fields of digital humanities and textual studies are each in their own way also concerned with describing and preserving digital material. As I have demonstrated in this chapter, the current
method of archival description in Canada is lacking in several key areas. This can be improved by collaborating with scholars in the digital humanities and textual studies fields; they have many of the same concerns as archivists, but approach their material from different historical, theoretical, and practical backgrounds. By working together, we can create a digital archives system that meets the needs of researchers while allowing archives to better manage their holdings. We can develop a way to describe born- and made-digital archival material that does the material justice.
Chapter Two: Digitizing the Analogue Archives

In 2015, at an event held at the University of British Columbia, the local student chapter of the Association of Canadian Archivists tweeted that Librarian and Archivist of Canada Guy Berthiaume asked how, as humanists, we translate the wonder of the original when we digitize our records.71 This crucial question is one that has been asked by digital humanists for decades, and one with which archivists must also engage. Indeed, there is even more than the “wonder of the original” involved in working with original analogue documents: as discussed in the previous chapter, the object’s materiality contains information about its creation and provenance that cannot easily be conveyed in an image. Archivists must be mindful of the losses incurred in the digitization process and find ways to convey this information in the metadata surrounding the digital object. There are advantages and disadvantages to digitization, but it is clear that, at this point, there is no turning back from it. Making archival material available online means that archives and their material can stay relevant, but archivists must think carefully about their long-term strategies and objectives for digitization and description before they embark on a digital archives project.

The Digitization Process

Before delving too deeply into the consequences of digitization, it will be helpful to explain what exactly is meant by that process. Digitization is the process of converting something analogue, like text on a piece of paper, to something digital, like an image of that piece of paper on a screen. It is the translation of something tangible and human-readable into something machine-readable, the transmission of words into a series of zeroes and ones. For

71 ACA UBC, Twitter post, February 13, 2015, 12:18 pm, https://twitter.com/ACAUBC/status/566300390312775680
most archives, and most archival material, this is done by using a scanner and following
standards created by regional, national, or international archival organizations.

In Canada, there is not one nationally-accepted standard for digitization. The Canadian
Council of Archives (CCA) links to a variety of resources on its website, including handbooks
from the US, UK, and Australian governments, as well as from educational institutions in a
variety of countries. While it is helpful to have these resources rounded up in one location, the
multiplicity of standards to choose from can make it difficult for archivists, particularly those
working in small institutions where they may be the only staff member, to choose the best
standard to make their digitized material stable and accessible for as long as possible.

Many archives in Canada follow the standards set out by the United States Library of
Congress or Cornell University. In their helpful online resource *Moving Theory into
Practice*, the Cornell University Library has a chart of the various digitization requirements
from different institutions in the US. Most of these institutions have different requirements for
different kinds of material—printed text tends to be at lower resolution than photographs, for

---

72 The Canadian Council of Archives offers a PDF of a list of links to other resources; however, this list has not been
updated since 2011. Both its format (a PDF containing links, rather than a webpage that would be easier to access)
and its age (leading to both out-of-date information and dead links), as well as the lack of any cohesive standards
regarding digitization in Canada, demonstrate the haphazard nature of digital archiving in Canada. Canadian Council
of Archives, “Step-By-Step Guides to Digitisation Projects,” accessed June 8, 2017,


74 Cornell University Library, *Moving Theory into Practice: Digital Imaging Tutorial*, accessed May 18, 2017,
http://preservationtutorial.library.cornell.edu/.

75 This resource is no longer being updated, but remains online as a reference tool. While the data referenced in
footnote 6 may be out of date, they still serve as a record of the differing digitization requirements that were used in
the early 21st century and illustrate the lack of consensus regarding those requirements.

76 Cornell University Library, “Table: Representative institutional requirements for conversion,” *Moving Theory into
Practice: Digital Imaging Tutorial*, accessed May 18, 2017,
https://www.library.cornell.edu/preservation/tutorial/conversion/table3-1.html.
instance. The Northeast Document Conservation Center also offers an online course called Preservation 101, which contains a section on digitization and digital preservation.\textsuperscript{77} The most authoritative and recent standards for archival digitization, however, are in \textit{Technical Guidelines for Digitizing Cultural Heritage Materials}, created by the Still Image Working Group at the American Federal Agencies Digitization Guidelines Initiative.\textsuperscript{78}

One of the most important aspects of digitization is resolution. It is crucial to take this into account in order to create a digital facsimile of the archival material that represents, as accurately as possible, the analogue original. Resolution, often expressed in dots per inch (DPI) or pixels per inch (PPI), refers to the number of pixels and how far apart they are spaced. A higher number means more information is packed into a certain amount of space (i.e. the image is of higher resolution) and a lower number means that there is less information in the same space (i.e. the image is of lower resolution). In general, the standard for digitizing material larger than a photographic slide is a colour scan (even for black-and-white or greyscale material) at 600 dpi, saved as a TIFF (Tagged Image File Format). The resulting large, archival-quality files can then be used to create smaller, more easily shared JPEG (Joint Photographic Experts Group) files.

The CCA does provide its users with a thorough guide to the process of deciding to digitize archival material.\textsuperscript{79} This document is relatively old, from 2002, but the list of “Principles Concerning the Relationship of Digitization to Preservation of Archival Records” it contains


remains important today. Particularly relevant to the topic of this thesis is the fifth principle:

“Digitization must strive to preserve to the greatest extent possible the authenticity and integrity
of the original information.” 80 This must be done not only through the ways the original archival
material is digitized, but also how it is described, which is also brought up in the CCA’s list of
principles: “9. Search tools are an essential part of a digitization project and must meet the needs
of users.” 81 Archival search tools rely on metadata—information about the material being
digitized and about the process of digitization—for the information that can be searched by
archival users, allowing the material to be discovered and used.

Metadata

An archival document is much more than a collection of words on a page. Joan M.
Schwartz and Terry Cook suggest that

the individual document is not just a bearer of historical content, but also a reflection of
the needs and desires of its creator, the purpose(s) for its creation, the audience(s)
viewing the record, the broader legal, technical, organizational, social, and cultural-
intellectual contexts in which the creator and audience operated and in which the
document is made meaningful, and the initial intervention and on-going mediation of
archivists. 82

All this information Cook and Schwartz describe, everything that is not “historical content”, can
and should be contained in metadata so that archivists and users of archives are clear that the
contexts of the records with which they work are more complicated and more interesting than
they may initially appear.

---

80 Ibid., 1.
81 Ibid.
In addition to providing helpful search terms for users, metadata also contains important information about the provenance of the digital item itself. Indeed, all the information contained in a traditional finding aid—all the information that \textit{RAD}, for instance, calls for—is metadata. This data about data helps us to contextualize the analogue archival artifact in the same way that it helps us to contextualize the digital artifact: it turns something that is only readable on a very surface level into something that \textit{means} something, something that has a history and a purpose. Metadata, then, informs everything we know about and how we interact with the archival object.

When it comes to digitized or born-digital material, however, metadata becomes even more important to our understanding of the archival object. Because digital material at its essence is machine-readable, not human-readable, we rely on the information we surround the data with—the metadata—to be able to understand what we are looking at. Even if a digital image of the item itself is visible to us, the lack of contextual information attached to a digital object compared to a more tangible analogue one means that we need the additional information contained in metadata in order to better understand the object. There are a number of metadata standards that exist; in Canada, many archives use \textit{RAD} to delimit the metadata they include in their descriptions.

This information about archival material—its creator, its date of creation, keywords or subject areas, etc.—helps both to identify and to provide context for the material. Metadata is also often structured in a way that makes it easy to create and to use. There are other forms of metadata beyond that used in most archival descriptions, of course—automatically generated information about the brand of camera used to take a certain photograph, for example. With all the different information that can be part of an object’s metadata, it follows that there are different broad, and not mutually exclusive, types of metadata standards for different groups of
users. Because there is so much information that can be conveyed about archival material, it is important to organize this metadata in ways that do not overwhelm the users but do allow them to find the information they need. Metadata standards mediate this information so that researchers can do their work without drowning in the flood of data that can be generated from each piece of material.

Different groups create metadata schemata based on what information they think it is important to preserve, and organize them based on their own needs and usage patterns. There are also schemata based around media types; that is, a certain way of dealing with metadata for digital audio files, and another for certain clerical documents. The lack of a cohesive system of dealing with various types of files means that these media-specific schemata are not very useful for institutions that contain more than one kind of material, like most archives. Overall, there are many different kinds of metadata standards out there, none more intrinsically valid than any other, but all devised to suit different needs. Archives must choose to which standards they plan to adhere, and make that choice—and the implications of it—clear to users of their collections.

To deal with rare books and manuscripts within an archival context, archivists must decide what information to convey about our material, and how best to arrange that information so that users can find out what they need as easily and with as little confusion as possible.

**Metadata Types**

The American National Information Standards Organization (NISO) divides metadata into three types: descriptive, structural, and administrative metadata.\(^3\) Administrative metadata can be further divided, into rights management and preservation metadata.\(^4\)

---


\(^4\) Ibid.
Descriptive metadata deals with the physical descriptive properties of the object and allows archivists and users to identify and find it. Examples of this include the size and quantity of the material, the names of people within a photograph, and the kind of script used in a manuscript. This information is often used in archival description to give researchers an idea of whether the object is what they are looking for before archives staff pull it from storage, and in textual description to inform researchers of the shape, size, and history of the text to which the description belongs.

Structural metadata gives context to how an archival object is constructed. For example, it allows a researcher to know how many pages a PDF file contains and how the information within the file is related to itself. This information is also related to codicology and the ways books and analogue material are structured—for example, a digitized version of a book comprises many discrete digital objects: images of each page, images of the book’s bindings, and descriptions of the book and each page within it. Without structural metadata, the relationships between these digital objects would not be clear, and the information contained in the objects would lose its meaning.

Rights management metadata contains information about the intellectual property and copyright on the material, and about who owns which rights to the material. This information is important if, for instance, an archives wants to put its digital material online. Reproducing archival material to which an archives does not own the rights can break copyright and intellectual property laws and leave archives open to litigation.

Preservation metadata deals with the information that archivists need to preserve the material to which it refers. This metadata also provides a record of any preservation measures taken by the archives. Examples of this kind of information include the file type and size, the
program that created the object, and the dates of the object’s creation and modifications. This information is crucial to knowing the history of the object, and particularly to knowing how to open and read the object. File formats can become obsolete very quickly, so it is important to know that a particular file is, for example, a WordPerfect document so that it can be preserved and used properly. Being aware of an item’s file type also allows archivists to plan for its preservation over time, and its possible migration to a more stable filetype. The PREMIS Event entity contains information about any actions taken that modify objects. These actions may include migration, normalization, and replication, as well as any other activities that change the object in order to preserve it.85

It may also be useful for archives and archivists to arrange the kinds of metadata according to the use they will have. Different communities of users have different needs: for example, a person looking for a book on a shelf at their local library can go to the catalogue and look up the book by its title or author. The catalogue will tell them the location of the book on the shelf, and they can go and retrieve it. The metadata included in the catalogue description is entirely sufficient for this kind of discovery use. However, if someone wants to know more information—like the size of the book, for instance, or the history of its existence and possession before it came to the library—they will likely have to look elsewhere for that information. As discussed in chapter one, a major shortfall of the current Canadian archival description standard is that, because it was originally based on a library cataloguing standard, it privileges information relevant only to published material over information that is crucial to understanding archival

material. Archival objects, particularly digital archival objects, need better description and preservation systems.

A major difference between born- and made-digital objects is that there is the chance to create a system to handle and preserve the made-digital objects before those objects are created. While organizations and individuals still largely do not have a way to preserve the last 20 years of emails and documents generated through day-to-day business, the act of digitizing something is a deliberate choice, one that has presumably been thought through. Because it is easier to handle digital objects born, so to say, into a structured system, made-digital items tend to be better organized and described than born-digital ones.  

**Metadata Standards**

In North America, many archives rely on metadata standards developed and maintained by the American Library of Congress. These standards include the Metadata Object Description Schema (MODS), used as a descriptive standard; the Metadata Encoding and Transmission Standard (METS), which defines the way structural metadata is arranged; and PREMIS (Preservation Metadata: Implementation Standards), which deals with preservation metadata.

MODS is a schema that uses the Extensible Markup Language (XML) to define descriptive elements of an archival or bibliographic object. It is based on the library community’s Machine-Readable Cataloging (MARC) format, and contains some of the same elements as MARC. No elements are mandatory in MODS records, though most records include a unique identifier, information about the type of resource, and the way it is encoded. Although

---

86In this thesis I am using “made-digital” as distinct from “born-digital” to distinguish between those objects that represent analogue items and those that do not; of course all digital photographs or scans, e.g., are also born-digital.
MODS is based on MARC, any system of cataloguing rules can be used with MODS, which makes it useful for archives as well as libraries.\(^87\)

METS also makes use of XML. It was developed in 2001 to provide “the means to convey the metadata necessary for both the management of digital objects within a repository and the exchange of such objects between repositories (or between repositories and their users).”\(^88\) In other words, METS does not define the content of a description, but rather serves as a mechanism to record the relationships among pieces of content, and between content and its related metadata. MODS records can easily be embedded within METS documents, as can other forms of descriptive metadata.

PREMIS is similar to MODS in that it provides a series of fields (referred to as “semantic units” in the PREMIS data dictionary), analogous to the MODS elements, that can be applied to each item. PREMIS does not define how metadata should be included in a record, but rather defines what kinds of information should be included. The PREMIS Data Dictionary defines preservation metadata as “the information a repository uses to support the digital preservation process.”\(^89\) Priscilla Caplan’s *Understanding PREMIS* notes that metadata can be used for preservation activities in the following ways: checksums can be used to ensure the authenticity of a file; information regarding the age and type of media can be included in metadata; metadata about file formats and their related hardware and software environments is necessary to migrate


or emulate the files; and the history of an item’s preservation actions can be recorded in metadata.90

Presentation

The physicality of the archival object is crucial to understanding it. Tom Nesmith argues that “the medium of a record [is] an aspect of its provenance…. The record originates in a medium used to make it, as well as in the ideas and purposes of its inscribers.”91 Because it is so difficult to portray information about a three-dimensional object accurately in a two-dimensional space, it is important to think carefully about the way digitized material is displayed. Particularly when preparing digitized material to be displayed on the internet, archivists must take care to provide as much context for the item, through the use of metadata, as possible. The materiality of the archival object cannot be transmitted, but information about its materiality—its size, physical condition, what it is made of, etc.—can be.

This is the kind of descriptive work that textual scholars have been doing for decades. By familiarizing themselves with the entire production process of, for example, the medieval book, textual studies experts know what distinguishes one book from another and what information is important to include. Descriptive bibliography was developed in order to convey this information to people who could not access the books it describes; for this reason, it is helpful to look at its process when considering how archivists could do a similar thing with archival description. As laid out by D.C. Greetham in his book Textual Scholarship: An Introduction, there are thirteen


aspects of a manuscript book that should be included in a complete descriptive bibliography.

These are:

1. The title of the manuscript….
2. The number of folios….
3. The material (papyrus, parchment, vellum, paper, including any watermarks if on paper. Any change in material should be given.
4. The general condition of the material.
5. The folio size and the size of the writing block….
6. The number of lines per writing block….
7. The date and type of binding, with a description of all tooling and other decoration.
8. The ‘collation’ (i.e. the makeup of the gatherings or quires…) [including any excisions or additions].
9. The contents, with folio numbers and any omissions or other peculiarities. … If any of the text of this particular manuscript has been printed (not just the text of the work, which may be extant in several different manuscripts), it should be mentioned here.
10. Decoration, describing typical floreations or historiated initials, and so on. Coats of arms, miniatures, and illuminations in general should be noted.92
11. The probable date and place of writing, followed by the general style of script, with specific indicators (characteristic letter-forms especially).
12. Evidence of the manuscript’s history from coats of arms, signatures of ownership, sales catalogues, catalogues of libraries, etc.
13. Printed notices, including descriptions or other references to the manuscript in bibliographical works.93

Of these components, the number and size of folios, material (and its condition), binding, and collation are the most crucial to understanding the physical attributes of the text itself. The number of folios gives an idea of the size of the book, while information about the material the text is written on can give textual scholars a look into the process of making the specific codex they are working with. Greetham points out that, as part of the processing of animal skins into parchment or vellum,

the skin was washed thoroughly, soaked in brine or lime, and dehaired, then stretched on a frame to be scraped, rubbed, and polished…. Although the actual hairs of the animal

92 A floreation or floriation is a floral decoration, often seen in the margins of medieval manuscripts; historiated initials are decorated with images of people or animals.

could be removed, the follicles remained, and this gave a peculiarly spotted appearance to the surface of the ‘hair’ side. In order to keep the ‘opening’ (a spread of two opening pages) as aesthetically consistent as possible, leaves were arranged so that the two ‘hair’ sides would always face each other, as would the two ‘flesh’ sides. Obviously, bibliographical ‘disturbances’ in a book—where the scholar suspects that the physical makeup has been changed—can often be detected by inconsistencies in this practice.94

In cases where the person interested in the item cannot see whether a page is the hair or flesh side of the leaf, such as when they are not with the item or if the scan of the page does not reveal this information, noting this information in the description of the item is very helpful. In addition to this, many digitization projects display one leaf at a time, rather than an opening of two leaves: any visual material information about the opening as a whole is lost.

Over the centuries that parchment has been produced, techniques for its processing have changed. Indeed, Greetham notes that the earliest parchment was probably somewhat coarser than that produced by the third or fourth centuries, when the new material had settled in as the dominant material. In later periods it varied significantly in quality, though certain general regional distinctions can be noted.95

These aspects of parchment’s texture can provide important information about the date and original location of the manuscript, but they may not come across in a photographic representation of the item. By including them in the description, textual scholars and archivists alike can ensure that researchers know as much as possible about the work they are studying.

An example of the importance of the materiality of records can be found in Ala Rekrut’s 2003 article, “Material Literacy: Reading Records as Material Culture.”96 In this article, Rekrut, head of conservation services at the Archives of Manitoba, discusses a letter from Métis

94 Ibid., 62.
95 Ibid., 62-63.
revolutionary Louis Riel to his wife Marguerite, written while he awaited his execution for treason. Rekrut examines the letter’s paper, pen, ink, and levels of deterioration, among other aspects of the text’s material form, to discern more information about the letter than is available from reading a transcription. As she points out, “[t]he physical characteristics of records provide evidence which may also be read in context to support or undermine the purported truth or authenticity of the text.”97 This evidence may be lost in a digital facsimile of the text and, as this thesis argues and Rekrut also points out, “current archival practices do not appear to support the systematic examination of physical evidence in records as a primary source of information, or the documentation of the evidence and the current understood significance of this evidence.”98

When it comes to books, bindings can contain a wealth of this evidence. Rekrut discusses the information that can be gleaned from the bindings of, for example, a letterbook of correspondence from the office of Manitoba’s Provincial Secretary from the late nineteenth century.99 The book the correspondence was copied into has certain characteristics, including the word “Letterbook” tooled in gold on the spine, that imply the use and purpose of the book.100 The discussion of this bureaucratic binding in Rekrut’s article is in stark contrast to projects like Library and Archives Canada’s digitization of ledgers from Kingston Penitentiary.101 While the pages of the ledgers were indeed digitized and can be viewed online, the material context of the pages—their covers and bindings—has been lost. Furthermore, the work of arranging the

97 Ibid., 22.
98 Ibid., 36.
99 Ibid., 14.
100 Ibid.
digitized images in order by page number has been done by a local history blog, rather than by LAC—the descriptions in LAC’s database can be arranged by relevance, date, or title, but not by page number. Indeed, Rekrut points out that oversized bound records have routinely been rebound to make them easier to store, quoting a 1991 guide to managing records that states “[b]ound records that are too large to fit in the deepest drawers, or with deteriorated bindings, can be rebound in post-binding format…. The record is not altered at any time during this process.” That the physical format of the record could be changed so dramatically during the archiving process but the record not be considered to have been altered speaks to the way archives and archivists have often ignored the materiality of records.

When it comes to medieval books, binding processes varied from binder to binder—indeed, since these books were often collections of texts that interested the person who had the book bound, the binding process can reveal much about the provenance of the book. To begin with, medieval books were not necessarily bound with hard wooden covers, as many surviving books are, or bound at all—some were kept in loose folios, while others had soft vellum covers. As Raymond Clemens and Timothy Graham point out, the type of wood used in a cover can provide an important clue as to the provenance of a book: “[t]he wood used for the front and back boards would typically be oak in northern Europe, beech in southern Europe.”

---

102 Library and Archives Canada, “Archives Search,” accessed May 25, 2017, http://collectionscanada.gc.ca/lac-bac/results/arch.php?FormName=MIKAN+Items+Display&PageNum=1&SortSpec=score+desc&Language=eng&QueryParser=lar_mikan&Sources=mikan&Archives=&SearchIn_1=partof&SearchInText_1=4291922&Operator_1=AND&SearchIn_2=&SearchInText_2=&Operator_2=AND&SearchInText_3=&Media=&Level=&MaterialDateOperator=after&MaterialDate=&DigitalImages=&Source=&ResultCount=10&canInd=. While the items can be arranged by title, the title of each object refers to the names of the people on the page, not the page’s number; it is impossible to reconstruct the body of the book this way.


Identifying characteristics of this wood may not be clear from a photograph; including this information in the description of an item can help a researcher learn more about the history of the book.

Clemens and Graham go on to discuss the process of covering the wooden boards of a book’s cover with parchment, on the inside, and leather, on the outside. The parchment used to cover the interior of the board is called a pastedown, and while binders sometimes used clean, blank parchment for this, they more often employed used and discarded leaves—“either leaves that had been discarded because the scribe had made an error on them or leaves that had been taken from outdated and unwanted manuscripts.”

By following the lead of textual scholars and including information about the physicality of the item, including its size, parchment, and binding style, in their description of the digitized material, archivists can ensure that as much of the information about the materiality of the artifact is preserved as is possible. As discussed in chapter 1, the Rules for Archival Description simply do not have the capacity to include this information in a helpful and easy-to-use way. A new kind of description, one influenced by digital humanists and textual scholars and their expertise translating the experience and physicality of the original into new forms, is necessary. This new description must also draw on the strong Canadian archival tradition of acknowledging the actions of archivists and researchers as part of the history and provenance of the archival object. As Elizabeth Yakel points out, archivists should “begin to think less in terms of a single, definitive, static arrangement and description process, but rather in terms of continuous, relative, fluid arrangements and descriptions as on-going representational processes.”

A digital object

---

105 Ibid.

may be rendered anew each time it is accessed, but the analogue object is arguably as unique every time it is seen by someone new. Capturing the subtle changes made every time the archival item is read or (especially) redescribed is crucial to understanding the history and meaning of the item.

It is also important to have the description of the item connected to its digital image(s) in order to provide users with the context they need to interpret the item. Particularly in an online environment, where images can circulate freely without any attribution whatsoever, tying the description to the digitized archival object is crucial. There will always be those who disseminate historical and archival photographs carelessly and without attribution. This is particularly notorious on social networks like Twitter, where there is limited space for contextual information, but archivists must do our best to provide contextual information about the material we create and share. As archivist Mark Vajcner points out, digital archives projects answer some of the same accessibility questions as publishing collections of documents did and do, but “tend to be devised for a broader audience where contextual questions are not part of the methodology.”108 Archivists must not assume that users of archives have been trained to treat all archival material with a skeptical eye, but should rather educate users about the nature of archives and archival work so that it is clear why contextual information is crucial to understanding an object. Popularly used tools like Access to Memory (AtoM), developed by Vancouver’s Artefactual, and Flickr allow archival description to exist on the same page as an

107 See, for example, the popular Twitter feeds @HistoryInPics, @HistoricalPics, @History_Pics, and so on. These feeds share archival photographs without attribution and are intensely popular; see Alexis C. Madrigal, “The 2 Teenagers Who Run the Wildly Popular Twitter Feed @historyinpics,” The Atlantic, January 23, 2014, http://www.theatlantic.com/technology/archive/2014/01/the-2-teenagers-who-run-the-wildly-popular-twitter-feed-historyinpics/283291/.

image of the archival object. On one hand, these tools are not constrained by Twitter’s 140-character limit; on the other, they are not nearly as socially popular.

It is impossible, or at least irresponsible, to talk about digitization of analogue records without talking about the digital records that digitization creates. As I discussed above, digitization alone is not the solution for the preservation of analogue material to begin with: many aspects of the analogue cannot be replicated in digital representations of it. Digitization reduces records merely to the information that is coded into text, while the information intrinsic to the materiality of the object is lost. As Rekrut points out, “[t]he physical object is already a conceptual, as well as a physical, ‘data’ object.”

When we digitize analogue material, we need to have a plan in place to take care of the digital material we create. Digital material deteriorates much more quickly and is much less stable than analogue material. Mike Kastellec points out that two issues are “at the core of digital preservation: data loss and technological obsolescence.”

Digital media is at risk for the same kinds of physical decay as analogue media: over time, stored in an improper environment, the supporting medium starts to break down. However, where in paper or parchment books this may lead to worm holes or discolouration of the pages, the data loss caused by physical degradation of digital media, sometimes referred to as “bit rot,” can quickly lead to the data becoming altogether unreadable. Technological obsolescence is the lack of availability of hardware or software that is able to read the digital media and turn its bitstream of ones and zeroes into human-readable data. There are strategies that can be used to overcome these issues: migrating digital material forward to current

---


hardware and software systems, maintaining multiple copies of items to decrease the odds of an item suffering physical degradation, converting material to more stable formats, and simulating a digital object’s original environment in order to access it using older tools. Because no storage medium is immune to bit rot, archivists and other information professionals keep multiple copies of archival material on different media. Bit rot, when it occurs, is detected by running checksums on the data to ensure the copies are the same as each other; when one copy is corrupted, it is replaced by an uncorrupted one. This method, though effective at maintaining accurate copies of digital material over the long term, is incredibly resource-heavy and inefficient. As David Rosenthal points out, “[s]ociety’s ever-increasing demands for vast amounts of data to be kept for the future are not matched by suitably lavish funds. Thus, absent a technological miracle, bit preservation is a problem with which we are doomed to struggle indefinitely.”

This is not to argue that digital preservation projects are not worth undertaking; demands for all digital material to be available forever, however, are unrealistic. Archivists must plan for the cost of digital preservation when proposing digitization programs: undertaking a massive digitization project is useless if the data created will be inaccessible in a few decades. In order for the project of digitization to make sense as something for archives and archivists to put their time, energy, and resources into, digitized and digital material needs to be described and stored correctly itself. Matthew Kirschenbaum suggests “that the preservation of digital objects is logically inseparable from the act of their creation—the lag between creation and preservation collapses completely, since a digital object may only ever be said to be preserved if it is

---

111 Ibid.

accessible, and each individual access creates the object anew.”\textsuperscript{113} The act of translating the series of bits that comprises each digital object into something not machine- but human-readable constitutes a re-creation of the rendering of the analogue artifact and the re-representation of the digital artifact. Both of these artifacts have their own materialities, and archivists must represent both materialities in the archival description of the object(s).\textsuperscript{114} The analogue object’s materiality is crucial to understanding it and its place in the historical record, but so too is the materiality of the digital object crucial not just to understanding it, but also to preserving it and making it accessible to future researchers. Digitization is not a simple task: archivists undertaking a digitization project must think carefully about the ways they represent the physical aspects of an analogue object on-screen as well about the ways they describe and preserve the physical aspects of the digital objects they create.

These tactile aspects of interacting with books can also provide clues as to the book’s provenance and production. As Ala Rekrut points out, people consciously make choices about the material they use when creating things in order to convey meaning, but archivists may overlook the choices of other creators when working with archival material.\textsuperscript{115} The style and thickness of the paper, the kind of type or handwriting used, and the type of binding, if there is one, can all give us information about how the material was produced and how it might have


\textsuperscript{114} As Katherine Hayles, Matthew Kirschenbaum, and others have discussed, digital objects have their own materialities that affect the ways users read and understand them. Hayles argues that “material differences between media do matter, and matter significantly, if one wishes to account for the specificity of reading practices, the responses of users or readers to particular texts, and the nuanced effects that different kinds of texts can achieve.” Katherine Hayles, “What Cybertext Theory Can’t Do,” Electronic Book Review, February 15, 2001, accessed May 26, 2017, http://www.electronicbookreview.com/thread/electropoetics/ecumenical.

\textsuperscript{115} Rekrut, “Material Literacy.”
been valued and used by earlier readers. An example from the world of rare books and manuscripts is Kathryn M. Rudy’s examination of discolourations on medieval missals and prayer books. The pages of some of these books are damaged by repeated kissing, while the fingerprints and stains in the corners of other pages show which sections were read the most. This information may not be accurately captured in the digitization process, and the wear and tear of everyday use may not be either. Rekrut points out that “current archival practices do not appear to support the systematic examination of physical evidence in records as a primary source of information, or the documentation of the evidence and the current understood significance of this evidence.”¹¹⁶ The importance of this evidence means that it must be included in archival description so that the information it conveys is not lost.

Another crucial form of physical evidence is size. The size of the page is hard to discern and the location of a text within a larger book is lost when an item is viewed on a screen. These are just two examples of the texts’ physicality that can be lost when they are digitized and examined as a single page of text, rather as part of a larger work. The size of a text is crucial to understanding its creation and purpose: as an example, there are nearly endless variations in the size of Bibles, depending on whether they were official versions used by the Catholic church or illegal vernacular translations used in secret. If every text appears to be the same size online, their histories and usages can be lost. The contents and order of collections of texts also provide important context to individual works. In books from an age where they were made to order for specific people and specific reasons, the texts that come before and after the one a person is interested in relate to that text. If a single part of a book or collection is looked at online, its context within the book is lost, and so is an important part of its meaning. Textual scholars have

¹¹⁶ Ibid., 36.
considered ways to convey the physicality of textual objects for decades. Particularly, the documentary editions of texts produced by textual scholars aim “to reproduce a manuscript or printed text as a historical artifact.”¹¹⁷ Digital humanists, drawing on the work of textual scholars, have considered ways to convey the information collected in these editions online for years as well. One way they have devised to do this is another form of metadata: TEI, or the Textual Encoding Initiative. Like MODS and METS, TEI uses XML to encode information about data. In this case, both information about the text, such as its authorship and dates of creation, and structural information about the document, such as its sections, chapters, and physical attributes, are included in the metadata. By using TEI to encode information about the materiality of a text, digital humanists and archivists can convey some of that materiality to an audience that only interacts with the text digitally.

Among modules that encode paragraph breaks, verse lineation, and words in languages other than the primary language of the text, the <msDesc> module “contains a description of a single identifiable manuscript or other text-bearing object.”¹¹⁸ This module is particularly useful to textual scholars and archivists, as it may contain information about the physical description of the object, as well as additional information about the object’s history, contents, and identifier. The <msDesc> module also contains components that deal with the ontological aspects of manuscripts: <msPart> can encode “information about an originally distinct manuscript or part of a manuscript, which is now part of a composite manuscript,” while <msFrag> “contains information about a fragment of a scattered manuscript now held as a single unit or bound into a


larger manuscript.”\textsuperscript{119} The \texttt{<physDesc>} element, which is used to encode the physical
description of a text, can be subdivided into further categories:

- \texttt{<additions>:} any significant additions, such as marginalia or other annotations
- \texttt{<bindingDesc>:} the present and former bindings of the manuscript
- \texttt{<decoDesc>:} the decoration of a manuscript
- \texttt{<handDesc>:} the kinds of hands and writing forms used
- \texttt{<musicNotation>:} the nature of any musical notation
- \texttt{<objectDesc>:} the physical components making up the object
- \texttt{<sealDesc>:} the seals or other external items attached to a manuscript
- \texttt{<typeDesc>:} the typefaces or other aspects of the printing of an incunable or other
  printed source
- \texttt{<accMat>:} closely associated accompanying matter\textsuperscript{120}

The flexibility in physical description that TEI provides means it is ideal for (and indeed was
developed by) humanities scholars, but it can also be used to describe aspects of archival
material that RAD and other archival standards do not support.

TEI can also be used to create digital facsimile editions of texts: as James Cummings
points out, though digital images of manuscripts are a desired part of digital editions, “one must
also provide a carefully edited full text” in order to create a scholarly edition of a work.\textsuperscript{121} The
\texttt{<facsimile>} element of TEI can be used to do this through linking facsimile images with the
corresponding transcription of the text. \texttt{<zone>} elements can be used to delineate specific areas
of an image using x and y coordinates, allowing editors to map transcription to very specific
areas of an image. As Cummings points out, “[h]aving discrete textual locations (say on a word-
level granularity) marked and related to a digital surrogate enables different methodologies that

\textsuperscript{119} Text Encoding Initiative Consortium, “Manuscript Description,” \textit{P5: Guidelines for Electronic Text Encoding

\textsuperscript{120} James Cummings, “The Materiality of Markup and the Text Encoding Initiative,” in \textit{Digitizing Medieval and

\textsuperscript{121} Ibid., 68.
are only starting to be exploited.”

This might eventually allow OCR of manuscripts, if recurring words or characters can be recognized, but in the meantime allows for increased clarity when it comes to reading and understanding manuscripts.

The sustainability of digitization and digital exhibition projects is also something that archivists must consider. While institutions often receive funding for the initial digitization work, the ongoing personnel and financial costs of keeping the digitized archive functional, accessible, and secure are often not funded as generously. Without certain institutional support for digital archives, it is hard to see a way forward. As discussed above, long-term digital preservation and storage is immensely costly. Adequate project funding is essential when planning any digitization project. While it is, of course, possible for institutions to charge the public for access to their material, this is at odds with the archival goals of accessibility and inclusivity. Increased funding from parent institutions and public sources is one way to deal with the financial constraints of digitization projects. Tom Evens and Laurence Hauttekeete point out that

[as digital preservation conserves cultural memory and ensures permanent access to information, cultural heritage institutions should be high on the political agenda…. ]he government has an important role to play in opening up archives and optimizing their accessibility…. Indeed, investments in digital heritage would allow institutions to engage audiences and develop mutual relationships with other stakeholders. But even more important, increased accessibility would allow socially vulnerable groups…to fully participate in this richness of content.\footnote{\textsuperscript{123}}

It is true that archives may be able to attract outside stakeholders in the process of digitizing and disseminating their material online. Elizabeth Yakel points out the fact that archives are being pushed “toward more open access… [at the same time as] they are being squeezed from the other

\textsuperscript{122} Ibid., 71.

side by commercial ventures, such as Ancestry and Footnote, which are providing enhanced access to archival materials for a fee.”

Yakel states that consumers seem willing to pay the fee, but this is not the case for all users. The Association of Canadian Archivists states on their website that they work “to ensure…the preservation and accessibility of Canada’s information resources and its documentary heritage.” If those resources and that heritage are behind a paywall, users of archives are not guaranteed access to them. Institutions and the government must fund digital preservation projects properly if they want to see them succeed.

---


Case Study

In the case of the University of Manitoba fragments, digitization may prove crucial to fully understanding their significance. Including the provenancial and physical information that can be included in TEI descriptions will help users of the archives learn as much as they can from the digital images of the fragments so that they can contribute anything they know to the description. Because the fragments have been removed from their material wholes, it is important to include information about their materiality in their description to help situate them within those wholes. Unlike the Ege fragments discussed in chapter one, it is impossible to rejoin these fragments with other pieces of the books from which they came. However, including material characteristics in the description will help the reader to understand the origins of the fragments and how they came to be where they are.

The fragments at the University of Manitoba have been preserved for two reasons: they are either being used as binding supports or as art. Four of the fragments are hanging on the wall of the St. John’s College Library, while one is being preserved at the School of Art—these fragments were removed from their home books with care, and were preserved as exemplars of artwork, handwriting, or texts. The binding support fragments, on the other hand, were preserved primarily for their strength and size, rather than for their beauty or contents (there are, however, two examples where the fragments used as covers for books were likely chosen to provide a striking look for the books). It is important to convey the materialities of both these kinds of fragments, and discuss the characteristics of each. As David Watt and I discuss in a forthcoming article, “fragments as art objects focus our attention on the ways that scribes and artists made the
Since these fragments were preserved for their beauty, it is crucial to include descriptive and contextual information about the material beauty of the items—and the work of these scribes and artists—in their descriptions. Similarly, the fragments used as binding fragments have material attributes that ought to be described: the handwriting and parchment used can provide important information about the provenance of the fragments. A particularly interesting example of a binding support fragment is found in Dysart 11 in the University of Manitoba Archives & Special Collections’s Dysart Collection. This fragment, wide enough to cover the entirety of a book, was originally part of a Torah scroll, and was likely chosen because as a scroll it was blank on one side, perfect for the cover of a book. However, the use of a Torah in this way is highly unlikely to have been sanctioned by a member of the Jewish community—when Jewish sacred texts, including the Torah, reach the end of their lives, they are usually retired and buried in a cemetery. In the case of the Dysart 11 fragment, the materiality of the artifact is what led to its selection as a binding support, but it also gives researchers a clue into the provenance of the book it covers and what kind of events may have occurred to destroy this sacred text in this way. Conveying this materiality in online description is crucial to understanding not just the fragment, but the book it covers, and the town it is from. The material aspects of the fragment could be conveyed in the UMASC description of the fragment, if there were more optional fields to describe the fragment than currently exist.

The components of description discussed by Greetham above can be mapped to TEI fields and included in online descriptions. The <msDesc> element and its subelements were

---

126 David Watt and Elizabeth-Anne Johnson, “Medieval and Early-Modern Manuscript Fragments at the University of Manitoba,” (unpublished draft), 13.

127 Information about the lifecycle of a Torah was provided by Justin Jaron Lewis.
created to enable this kind of work; there is even an `<msFrag>` element that is designed specifically to contain information about “a fragment of a scattered manuscript now held as a single unit or bound into a larger manuscript.”128 The title of the manuscript, the number of folios, the material and condition of the material, and so on: all these aspects of the material fragment can be included in a TEI-compliant description, in the `<msIdentifier>`, `<collation>`, and `<extent>` fields, respectively. By incorporating Greetham’s components of description, through TEI, into archival descriptions, we can provide a far more detailed, accurate, and evocative view of the medieval manuscript fragments at the University of Manitoba than either library cataloguing or archival description would.

**Conclusion**

Original archival materials can be, as Guy Berthiaume points out, wondrous. While archivists may not be able to translate all of their wonder onto servers and electronic screens, they can and must do their best to make archival material accessible to users outside of their reading rooms. This digitization process must include the creation of finding aids that encompass all the information archivists and users need to understand the original material and the steps archivists took to make it available digitally. Incorporating the kind of work on materiality done by textual scholars into archival description by adding data according to the TEI guidelines will allow archivists to better describe the made-digital material in their holdings. The metadata included in archival descriptions needs to be clearer about the material itself and about the processes involved in selecting, describing, and preserving the material. By reaching out to other fields of study, archivists can create this kind of description.

---

Chapter Three: Crowdsourcing Archival Description

“When you see those pictures, it’s like coming home,” said Inuit Elder Piita Irniq at the fifteenth anniversary celebration of Project Naming, a Library and Archives Canada crowdsourcing initiative. Project Naming was started by Murray Angus, an instructor at Nunavut Sivuniksavut, a college in Ottawa. As Beth Greenhorn, the project manager of Project Naming, explains, Angus and his colleagues “have been bringing their students to Library and Archives Canada to do research in the card catalogues in our reference area to look for photos of their community and sometimes family pictures that they could take home at Christmas time.”

Because there can be a large disconnect between contemporary Inuit youth and their elders, history and culture, these photographs were a way to bridge the gap:

Greenhorn: The project was proposed by Murray largely because most of the photographs of Inuit in our collection were never identified. A lot of these photographs date from the early 1900s to the 1920s and even as late as the 1960s and 1970s. This was a way, as Murray saw it, to reunite and bring together two generations within Inuit communities. A large majority of Inuit youth today do not speak Inuktitut, which is the first language of Inuit; whereas the older generation doesn't speak English, so there really has been this generational gap. Another reason for suggesting this project is that, besides a loss of language, many youth don't have knowledge or an understanding of their past; it's not taught in their curriculum. By looking at these historical photographs, not only is the younger generation able to talk to the older generation, but it is also a way for the younger generations of Inuit to learn about their past. So really, it has been a way to bridge together these two generations.

From these beginnings, the impact and scope of the project have expanded to engage Indigenous people across Canada and not only gather the subjects’ names, but update the outdated and


131 Ibid.
inaccurate information sometimes included in the photos’ original descriptions. In this case, a crowdsourcing project has created engagement between an archives and its users that goes far beyond simply collecting data; indeed, the data collection is a secondary benefit to the connections that participants can form with their history and communities.

Crowdsourcing, a term coined in 2006 by Jeff Howe in an article for *Wired*, has been defined several ways. Howe based its etymology on “outsourcing,” the process of contracting external workers or companies to provide services for an organization at a reduced rate. In his article, Howe focusses on corporate uses of crowdsourcing such as stock photography provider iStockphoto and research hub InnoCentive. As indicated by Enrique Estellés-Arolas and Fernando Gonzáles-Ladrón-de-Guevara’s comprehensive description, however, it is clear that crowdsourcing can benefit many more sectors than just the corporate world:

Crowdsourcing is a type of participative online activity in which an individual, an institution, a non-profit organization, or [a] company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. The undertaking of the task, of variable complexity and modularity, and in which the crowd should participate bringing their work, money, knowledge, and/or experience, always entails mutual benefit. The user will receive the satisfaction of a given type of need, be it economic, social recognition, self-esteem, or the development of individual skills, while the crowdsourcer will obtain and use to their advantage what the user has brought to the venture, whose form will depend on the type of activity undertaken.

While crowdsourcing can be used for a wide variety of projects, there have been a number of notable successes in the field of cultural heritage. In his book *Crowdsourcing*, Daren C. Brabham discusses the different kinds of tasks crowdsourcing can be used for and divides them into four

---


groups: “knowledge discovery and management, broadcast search, peer-vetted creative production, and distributed human-intelligence tasking.”¹³⁴ The majority of crowdsourced work sought by galleries, libraries, archives, and museums (GLAM) consists of the final kind: this “approach to crowdsourcing is appropriate when a corpus of data is known and the problem is not to produce designs, find information, or develop solutions but process data.”¹³⁵ Because heritage institutions have so much material that is not easily machine-readable, recruiting people to help interpret this material at a very basic level allows the trained professionals who work in these fields to deal with more complicated tasks. For example, Library and Archives Canada’s groundbreaking “Project Naming” project, discussed above, seeks to identify Indigenous people pictured in LAC’s photographic collections. In its early stages, this project digitized photos of unidentified Inuit people and transferred them to laptop computers which could then be taken to the communities where the photographs were taken.¹³⁶ “Project Naming” and its success are excellent reminders that digital archival work need not always be online archival work. The American and Australian national archives have also been leaders when it comes to crowdsourcing: the American National Archives and Records Administration’s “Citizen Archivist” project asks volunteers to transcribe documents, tag records, subtitle videos, and edit Wiki pages.¹³⁷ It has been in operation since 2012 and was identified as one of the Top 25 Innovations in Government by Harvard’s Ash Center for Democratic Governance and


¹³⁵ Ibid., 50.


Innovation. The Citizen Archivist project has been a success—in response to a transcription challenge that took place over one week in 2015, NARA volunteers added 10,000 tags to records and transcribed 2500 pages of archival material. The Australian government also relies on crowdsourcing to make its records more accessible: the National Archives of Australia also launched “The Hive” in 2012. This project focusses on the transcription of handwritten or typed file lists, in order to make them searchable and the material they list more accessible to the public. Both of these national crowdsourcing projects use the power of the crowd to process data into a machine-readable format, allowing trained archivists to focus on more challenging and technical work.

Crowdsourcing and Cultural Heritage

One of the ways crowdsourcing can be of use to cultural heritage organizations in general and archives in particular is to allow users to contribute information about the cultural artifacts held by the organizations. In addition to increasing the amount and kinds of information available to other users of archives, the crowdsourcing process problematizes the common ideas that either records or archivists can be impartial. These ideas are rooted in the history of archival theory, particularly in the work of British archivist Hilary Jenkinson.

Jenkinson wrote in his *Manual of Archive Administration*, first published in 1922, that

> A document which may be said to belong to the class of Archives is one which was drawn up or used in the course of an administrative or executive transaction (whether public or private) of which itself formed a part; and subsequently preserved in their own custody for their own information by the person or persons responsible for that transaction and their legitimate successors.

---

138 Ibid.

139 Ibid.

To this Definition we may add a corollary. Archives were not drawn up in the interest or for the information of Posterity.\footnote{Hilary Jenkinson, \textit{A Manual of Archive Administration} (London: Percy Lund, Humphries and Co., Ltd., 1937), quoted in Theodore C. Pease, “A Manual of Archive Administration by Hilary Jenkinson,” \textit{The American Archivist} 1, no. 1 (January 1938): 23.}

This definition of archival records excludes much material that many modern archivists would consider to have archival value. However, Jenkinson’s ideas have continued to be supported by some archival theorists. The proponents of what Mark A. Greene has dubbed the “recordkeeping paradigm” of archival theory and practice echo Jenkinson’s thoughts about the way archival material is generated during the transactions of people or institutions.\footnote{Mark A. Greene, “The Power of Meaning: The Archival Mission in the Postmodern Age,” \textit{The American Archivist} 65 (Spring/Summer 2002): 42-55.} Luciana Duranti argues that “archival material is impartial evidence of actions and transactions and provides a reliable account of them.”\footnote{Luciana Duranti, “Commentary,” \textit{The American Archivist} 57, no. 1 (Winter 1994): 36.} Whether or not archivists believe this to be true, describing archival material from a singular and authoritative point of view, as if the only way to convey the meaning of the record were to repeat how it was generated and its chain of custody, reinforces this paradigm of archival thinking. Records and their meanings are more complex than the recordkeeping paradigm allows; archival description must allow space for that complexity.

The other paradigm described by Greene is the “archival paradigm.” In contrast to the Jenkinsonian recordkeeping paradigm, this approach to archives defines records as

\begin{quote}
\ldots any type of recorded information, regardless of physical form or characteristics, created, received, or maintained by a person, institution, or organization…. Records are extensions of the human memory, purposefully created to record information, document transactions, communicate thoughts, substantiate claims, advance explanations, offer justifications, and provide lasting evidence of events.\footnote{Bruce Dearstyn, \textit{The Archival Enterprise: Modern Archival Principles, Practices, and Techniques} (Chicago and London: American Library Association, 1993), 1, quoted in Greene, “The Power of Meaning,” 44.}
\end{quote}
This broad scope allows for much more than the byproducts of transactions to be considered archival. The multiplicity of voices whose words and stories are now considered archivally valuable ought, in turn, to be represented as part of the ways archivists describe those stories.

Just as records’ provenance gives them a wealth of meanings before they even arrive at an archives, so too do the ways in which they are described give them different meanings once they have achieved “archival” status. The ways records are used and read, as well as their users and readers, also all affect their meanings. This variety of meanings means that a record simply cannot only be described in one way; as Greg Bak points out, “[a]s the meaning of a document shifts, so does its classification.”145 The shifting meanings of archival records mean that no static description can do them justice; indeed, the search for a monolithic classification system for records is a fool’s errand. Bak goes on to state that

There is no ‘natural classification’ in the sense of a classification system that reveals an unmediated order that would exist regardless of human perception—a notion that is particularly absurd when applied to documents, since documents do not exist in the absence of human beings.146

Documents are not sterile byproducts of the functions of people or corporations. They are created by human beings, admittedly for specific reasons, but not for or through those reasons alone. Creating a space for users of archives to contribute their own understandings of archival material will help to emphasize this point and demonstrate the complicated past—the intersections and interrelationships, the original creation and subsequent recreations by users—that all archival documents have.

146 Ibid.
Allowing users of archives to contribute to archival description will also allow them to understand the subjectivities of archivists and archives. To deem historical material as archival is to imbue it with an amount of authority and validity that other material may not have. Not all material is archival or has lasting value, but archivists must identify, or at least acknowledge, the choices they have made in shaping the historical narrative. Terry Cook reminds archivists that researchers only see a predefined and monolithic universe – predefined especially by the archivist. What they see is what they get. They do not see what archivists saw before the appraisal decisions were made to give researchers what they get, and they do not understand the underlying assumptions of how archivists have described what they are now seeing in descriptive tools that present the results of that appraisal and subsequent arrangements.\(^{147}\)

While it may seem daunting to document each and every step in the appraisal and arrangement process, there are simple steps archivists can take to provide greater transparency about their work and the decisions they make. For example, archivists can document their appraisal process and make that documentation publicly available through a link to the appraisal report in that collection’s finding aid. By sharing the thought processes behind what they consider archival and what they do not, archivists can make their own subjectivities clear. Acknowledging our own subjectivities as archivists also opens up a space for others to contribute their own knowledge on a topic or particular archival object. Terry Cook calls for a “fourth phase” of appraisal practice, one where “we engage our expertise with [citizens’] in a blend of coaching, mentoring, and partnering.”\(^{148}\) The partnership Cook envisions walks the line between fully closed archives that make none of their internal processes known and fully open crowdsourced projects that ignore


\(^{148}\) Terry Cook, “‘We Are What We Keep; We Keep What We Are’: Archival Appraisal Past, Present and Future,” Journal of the Society of Archivists 32:3 (2011): 182.
the knowledge of trained experts in favour of the opinions (and different expertises) of members of the public. Different projects approach this line differently, but where successful crowdsourced GLAM (galleries, libraries, archives, and museums) projects excel is in bringing the proficiencies of the public (whether gained through training or lived experience) together with the skills of trained information professionals.

**Transcribe Bentham**

One of the more notable crowdsourcing projects being organized by an academic institution is *Transcribe Bentham*, based at University College London. Transcribe Bentham seeks to transcribe the digitized work of eighteenth- and nineteenth-century philosopher and reformer Jeremy Bentham. Begun in September 2010, this project has succeeded in engaging the public’s attention and enthusiasm, with 18,355 manuscript pages being fully or partly transcribed by May 26, 2017. Transcribe Bentham is coordinated by UCL’s Bentham Project in partnership with UCL Centre for Digital Humanities, UCL Library Services, UCL Creative Media Services, and the University of London Computer Centre. The British Library, which holds an additional 12,500 manuscript pages of Bentham’s work, joined the project in October 2012.

Much of the work of Bentham, who is perhaps best known for his conceptual design for the panopticon prison, was unavailable to the public or to scholars in any form “which

---

149 “About Us.” *Transcribe Bentham*, accessed September 4, 2015, blogs.ucl.ac.uk/transcribe-bentham/


adequately represented his writings as he envisaged them." To combat this, and in order to produce a new critical edition of Bentham’s work, the Bentham Project at UCL had, by September 2015, digitized approximately 60,000 manuscript pages of Bentham’s work and imported them into the project’s Transcription Desk software. This software, a customized version of MediaWiki, allows contributors to the project to select a box and manuscript of Bentham’s work and begin to transcribe it. The Transcription Desk interface also allows contributors to mark up their transcription in TEI markup in order to accurately render a digital version of the original manuscript. Because users of the Transcription Desk cannot be expected to be familiar with TEI markup, the Transcribe Bentham installation of MediaWiki includes a customized extension that includes buttons, similar to those in a WYSIWYG text editor, which contributors can use to mark features of the text such as additions, strikes, illegible sections, and marginal notes. TEI-encoded texts can also be easily converted to a variety of other formats, which increases the longevity of the digital texts. Once a volunteer transcriber completes a text, they submit their transcription to a Transcribe Bentham staff member, who checks the accuracy of the transcription and the encoding. If completed, a text is locked from further editing; if there is still work to be done, the text remains available for others to work on.

The number of paid staff members at Transcribe Bentham has varied over the years, from two full-time research associates in the early years of its Arts and Humanities Research Council grant to only two days a week of staff time when that grant elapsed in spring of 2011. Despite the

152 Causer and Terras, “Many Hands Make Light Work….,” 59.


154 What You See Is What You Get, a form of editor that is commonly used to create and edit content in a form that closely resembles its final published appearance.
change in staffing, volunteers continued to work on the project, encouraged by the willingness of
Transcribe Bentham’s managers to ask for and listen to feedback on the transcribers’ concerns.
Transcribers are rewarded for their work with points: creating an account garners them 1000
points, as does uploading an avatar; adding a friend is worth 100 points, editing pages is 25
points, and adding comments is 10 points.155 There are ten levels to which transcribers can
ascend, based on the number of points they have. These levels are named in ways which reflect
the scribal basis of the project: some are apprentice, scribe, amanuensis, acolyte, and adept. The
usernames of the top 20 volunteers are listed on the main page of the project, along with their
avatars, point total, and level status.156 Engaging with volunteers in this way allows their work to
be acknowledged and a sense of community to be fostered—the Transcribe Bentham user
forums are currently under construction, but Martin Moyle and his co-authors note that the
project’s point system is “a lively topic of discussion on the forum” in their 2011 paper on the
project.157 The front page of the Transcribe Bentham Transcription Desk also includes a progress
bar that indicates how far the project has come: as of June 2017, the project is 44.41 per cent
complete, with 18,374 folia transcribed. A link below the bar leads to the “Benthamometer”, a
more detailed breakdown of the project’s progress. This transparency about the existing progress
of the project and about how far there is to go means that volunteers have an idea what they are
working towards. In particular, the breakdown of specific boxes’ contents on the detailed

bentham.da.ulcc.ac.uk/td/Help:User_levels

http://www.transcribe-bentham.da.ulcc.ac.uk/td/Transcribe_Bentham.

157 Martin Moyle, Justin Tonra, and Valerie Wallace, “Manuscript Transcription by Crowdsourcing: Transcribe
progress page gives a concrete sense of the physical pages the volunteers are transcribing and breaks the overall project down into more discrete sections.\textsuperscript{158}

In early 2011, \textit{Transcribe Bentham} organizers created a survey that asked participants what they liked most about contributing to the project, and what prevented them from transcribing more. From the results of this survey, the organizers learned that volunteers “took part mainly owing to interests in: Bentham’s life and thought; history and philosophy; crowdsourcing and the technology behind the project; and a sense of altruism, taking part in something which will ultimately benefit the wider community.”\textsuperscript{159} However, respondents to the survey also reported that their participation in the project was limited by “lack of time in which to learn how to transcribe Bentham’s handwriting; various issues with the Transcription Desk; the difficulty of deciphering Bentham’s hand; and the TEI mark-up was considered by several volunteers as an aggravation to an already demanding task.”\textsuperscript{160} Based on these survey responses, \textit{Transcribe Bentham} made changes to their transcription process: a new version of the Transcription Desk software incorporated new tools such as the ability to rotate the manuscript image and tabs which show, respectively, the transcription area, a preview of what the encoded text will look like, and a list of changes the transcriber has made to the text. Tim Causer and Melissa Terras, writing about their project in \textit{Crowdsourcing Our Digital Heritage}, state that “[f]eedback from volunteers suggests that they regard the updated website as cleaner, faster, and more inviting.”\textsuperscript{161} An improved user interface also ensures the quality of the transcription and the

\begin{flushleft}
\begin{footnotesize}
\textsuperscript{159} Causer and Terras, “Many Hands Make Light Work…,” 67. \\
\textsuperscript{160} Ibid. \textsuperscript{161} Ibid., 71.
\end{footnotesize}
\end{flushleft}
markup contributed by the volunteers: 40 per cent of the transcripts were approved by Transcribe Bentham staff members without any changes to the text.\(^{162}\)

Beginning in the third period of the project (1 October 2012-19 July 2013), when a full-time staff member was again working, the Transcribe Bentham team began keeping more detailed records of their use of time. On average, at this point, a manuscript page took 5 minutes 57 seconds to check.\(^{163}\) As Causer and Terras note, “it is a rare Bentham manuscript that we could transcribe—let alone encode—in that space of time.”\(^{164}\) However, it was still the encoding of the text of the manuscript into TEI that both took the most time to check and contained the most errors. Twenty-four per cent of the transcriptions needed 10 or more alterations to their markup; 45 per cent of the moderation time in the third period was spent on these transcripts.\(^{165}\)

A concern of crowdsourcing projects is the possible discrepancy between the number of participants and the amount of work they each do. In the case of Transcribe Bentham, the lion’s share of the work has been done by seventeen “Super Transcribers,” a fact which has led Causer and Terras to note that “Transcribe Bentham might be better described as ‘crowd-sifting’ [than as ‘crowdsourcing’]: that is, beginning with the traditional open call associated with crowdsourcing, and then encouraging the emergence of a self-selecting, smaller number of individuals with the skills, desire and time to complete a complex task on a regular basis.”\(^{166}\) It is true that a small group of people performs the majority of the work in most crowdsourcing

\(^{162}\) Ibid., 80.
\(^{163}\) Ibid., 77.
\(^{164}\) Ibid., 79-80.
\(^{165}\) Ibid., 81.
\(^{166}\) Ibid., 73-74.
projects, but the way Transcribe Bentham was gamified may also have limited the number of contributors: because the leaderboard lists the volunteers who have contributed the most over all time, rather than the top volunteers of each week or month, there is unlikely to be any movement in the chart. The large number of contributions made by experienced volunteers may also mean that newer volunteers feel intimidated or overwhelmed by the amount of work done by their predecessors.

This reliance on a small group of volunteers is a hallmark of crowdsourcing. As Daren C. Brabham points out, the majority of people who contribute to crowdsourcing projects are professionals in their own right:

> We must keep in mind that crowds are not, on their face, comprised mostly of amateurs. They are largely self-selected experts and what we might otherwise call professionals, who seek opportunities to make money, express themselves, build portfolios for future employment, and enjoy all the responsibilities and trappings of serious leisure.  

A person choosing to work for free does not mean they are not qualified or able to do similar work for pay. The tension between the common conception of the crowd as a source of unskilled (and therefore free) labour and the reality that the crowd that works on any given project tends to be educated and experienced in the field of the project must be addressed by those managing crowdsourcing projects. This brings up ethical questions regarding using free labour of people who might ordinarily get paid for the same work; as Brabham points out, “[d]istributed labor, whether outsourced overseas or crowdsourced over the Internet, is a hallmark of global capitalism and a proven strategy for deflating the power of unions and hindering labor organizing.”

---


168 Ibid., 405.
volition, their labour still has monetary value, as demonstrated by the small number of people who are able and willing to do the work.

**Ransom Center Fragments**

The *Harry Ransom Center Fragments*, a project started by then-archivist Micah Erwin in June 2012, is a project that attracted an especially specialized group of volunteers. This project sought to “survey […] and describ[e] […] the medieval and early modern leaf collection and manuscript binding waste in the Ransom’s book collection.” 169 Because Erwin was the Ransom Center’s only archivist with a background in medieval studies, he was put in charge of the fragment project; because his job description included much more than just the fragments, he reached out to others in the rare book and manuscript community for help. Erwin began posting images of notable medieval and early modern manuscript fragments on a Flickr account he created for the Harry Ransom Center fragments, and sharing each new addition on the project’s Facebook page. Erwin posted on the Ransom Center’s blog, *Cultural Compass*,170 and made use of the existing social networks in the medievalist community to quickly find a crowd to work on the project: 40 per cent of the fragments on the *Ransom Center Fragments* Flickr page were identified in the first few months of the project.171 Growth of the project was slow at first, but grew as more fragments were added and as the project was written about in blog posts and on the


Early Book Society listserv. The Harry Ransom Center fragments were viewed over 14,000 times by mid-October 2012, only four months after the project began.\textsuperscript{172} As Erwin reports in October 2012,

All images have been viewed at least once and 73 out of 225 images have comments. 21 of the texts on the fragments have been positively identified or at least attributed to print editions available online while another 13 fragments now include rough transcriptions or other relevant information.\textsuperscript{173}

The project, the name of which was changed to \textit{The Medieval Fragments Project} at some point, continued for over two years more, with Erwin sharing conference papers and presentations on his blog. The most recent of these presentations was posted on 1 June 2013, approximately a year after the project began; after this there was silence from Erwin until a now-deleted post on 25 May 2015. Entitled “Why the Medieval Fragments Project Nolonger [sic] Exists: and when crowdsourcing doesn’t work.,” the post discusses the termination of the \textit{Medieval Fragments Project} and Erwin’s unilateral decision to delete the project’s Flickr and Facebook pages in the wake of his contract at the Harry Ransom Center not being renewed. Erwin states that he made his decision because he “can no longer maintain any curatorial control over the fate of the fragments and because [he] cannot afford to expend any more time and effort managing the Fragments Project”.\textsuperscript{174} The result of the project, a description of its history as well as a catalogue of the fragments, has been available upon request from Erwin.

\textsuperscript{172} Erwin, “Crowdsourcing the Medieval Text.”

\textsuperscript{173} Ibid.

In his final blog post on the topic, Micah Erwin included three things that “advocates of crowdsourcing” should learn from his experience:

- Crowdsourcing initiatives should have the full and official support of an established institution if they are to survive in the long run.
- Crowdsourcing initiatives should never rely entirely on volunteer labor. The people who direct the project should have a financial stake in the enterprise.
- If a crowdsourcing project is to exist entirely on volunteer labor, then the subject of the initiative should not be under the curatorship of a cultural heritage institution.\(^\text{175}\)

Erwin makes the point that those involved in the administration of a crowdsourcing project should be compensated for their labour. Particularly in the cases of crowdsourcing work commissioned or planned by a funded and established heritage institution, it is crucial that the work of administrators, if not the crowd workers themselves, be compensated. Successful examples of community archives run by unpaid administrators exist, of course—the rukus! Black LGBT archive, before its transfer to the London Metropolitan Archives, and the crowdsourced Photos Normandie project among them—but in examples such as Erwin’s, it is important that institutions be financially engaged in their crowdsourcing projects. Initially paying someone to do work and then expecting them to do the same work for free is unrealistic and unsustainable. Again, as Brabham points out, the work of a crowd tends to be professional work (and in the case of the Ransom Center Fragments, it almost certainly was); even if an institution is happy to have the crowd do work, it ought at least to compensate the person or people in charge of organizing the project. Cultural heritage institutions—libraries, archives, museums—have a responsibility to the artefacts that are in their care, and the thrill of identifying and gathering new information about their material must be secondary to creating a sustainable online environment in which to display it.

\(^{175}\) Ibid.
The deletion of the *Ransom Center Fragments* raises interesting questions around the ultimate goals of crowdsourcing. While Erwin was able to complete a catalogue of 106 of the 121 manuscript fragments held at the Ransom Center, his deletion of the project’s Flickr page, which itself served as a kind of catalogue, means that other scholars and the community of contributors that made his project a success are no longer able to view the images of the fragments they worked on. In addition to this, no one is able to refer directly back to the images or their description.\(^{176}\) The photographs themselves were the intellectual property of the Ransom Center, but Erwin posted them on a Flickr account and Facebook page of which he seems to have been the only administrator.\(^{177}\) In addition to this, there are still fifteen unidentified fragments that will, presumably, remain unidentified. Micah Erwin should indeed have been properly compensated for his work on the fragments, and it seems likely that the Ransom Center did not understand or appreciate the value of the work he was doing. However, to undo all the work—hours of labour—contributed by volunteers simply out of spite with one’s ex-employer seems rash, ill-advised, and unethical. As Erwin suggests, those in charge of crowdsourcing projects should have a financial stake in its success. However, they should also have another—any other—stake in its success as well. Whether or not Erwin’s reasons for action were understandable, the action itself was irresponsible. The hands-off approach of the Harry Ransom Center when it came to such an important project is hard to understand—social media accounts were hardly new in 2012, when the *Fragments* project started. Leaving the reins of institutional


\(^{177}\) “Ransom Center Fragments” (https://www.flickr.com/photos/ransom_center_fragments/) and “The Fragments Project” (https://www.facebook.com/TheFragmentsProject). Neither website is extant or archived by the Internet Archives’s Wayback Machine.
Facebook and Flickr accounts in the hands of a single contract archivist was unwise, and the international manuscript studies community is poorer for that choice.

**Old Weather**

*Old Weather* is a crowdsourcing project that seeks to use ships’ logbooks to determine climatic data and changes in the weather over the world’s oceans. This project is a collaboration between multiple organizations and agencies, but was begun by Zooniverse, a citizen-science group; the Met Office; the National Maritime Museum; and Naval-History.net in 2010. The aim of *Old Weather* is to transcribe digitized copies of naval logbooks from ships in the British Royal Navy between 1914 and 1923 in order to create a database of oceanic weather data; with this information, climatologists and other scientists will have a better understanding of what past weather has been like over the oceans and be better able to predict future weather patterns.

*Old Weather* is rare among GLAM crowdsourcing projects because it is a scientific project; many other projects, including the others mentioned in detail in this chapter, deal strictly with humanities-based research. As Lucinda Blaser points out, “*Old Weather* was probably one of the first citizen science projects that combined common crowdsourcing projects, such as transcription, with a historical collection and a scientific goal.” She goes on to state, however, that the crowd the project attracted was more interested in the historical aspect of the project than she and the other project developers anticipated: “it became clear that it was the content of the

---

178 The entire list of collaborators, as found at http://www.oldweather.org/about, is: the National Archives and Records Administration, the National Archives of the United Kingdom, the UK Met Office, the American National Oceanic and Atmospheric Administration, the Pacific Marine Environmental Laboratory, the Office of National Marine Sanctuaries, the National Oceanographic Data Center, the National Maritime Museum, the New Bedford Whaling Museum, the Providence Public Library, the National Marine Sanctuary Foundation, the University of Oxford, the Citizen Science Alliance, the Atmospheric Circulation Reconstructions over the Earth Initiative, Naval-History.net, Jisc, and the Cooperative Institute for Research in Environmental Sciences.

documents—particularly daily events on board noted in the logs [rather than the weather observations]—that were keeping users coming back.” Indeed, the volunteer contributors to *Old Weather* became invested in the ships whose logs they were following, discussing their strategies behind choosing a new ship to follow on the project’s message boards.  

The relative sophistication and gamification of *Old Weather* compared to other crowdsourcing projects may have also had an impact on the level of user investment. Much like at *Transcribe Bentham*, volunteers are first assigned the rank of cadet when they sign up to the *Old Weather* site; by transcribing a certain number of logs, users can ascend through the “ranks” of the project to the level of captain. By creating incentives, even seemingly ephemeral ones, to keep contributing, *Old Weather* has developed a broad and invested userbase. The first two phases, revolving around the Royal Navy logs, were completed by September 2012, with 16,400 people transcribing over one million pages.  

**Lessons for archivists**  

Archivists can gain important perspective from looking at all three of these case studies. *Transcribe Bentham* is notable for its use of user-friendly transcription software and for incorporating user feedback into its process. Engaging with the crowd through surveys as well as through the *Transcribe Bentham* website allowed the managers of this project to adapt the Transcription Desk interface to better suit their users. As archivists, we need to reach out to our

---

180 Ibid.  
181 Ibid.  
users and find out what they enjoy about using our services and what we can do to improve their engagement.

Transcribe Bentham also deals well with the issue of the role of the professional in a crowdsourcing project. Although the crowd may do the work of transcribing, professionals are still responsible for checking and finalizing that transcription. Far from devaluing the work of trained professionals, this project puts them in a position of authority while still freeing them from the work of transcription. This project is also a good example of how to proceed with a crowdsourcing project once grant money has evaporated. Despite staffing shortages, the engagement already built up with the Transcribe Bentham crowd allowed the project to continue through more fallow funding periods and to ramp up again once more money was available. This is a situation which is likely familiar to many archivists; by planning a project well when funding is available and by creating a good relationship with our users, we may be able to continue aspects of our work even when extensive funding is not available.

Transcribe Bentham and Old Weather have also gamified their process by the inclusion of user levels based on users’ participation in the project. In the case of Transcribe Bentham, the names of the levels also provide a link to the past and allow the contributors to feel like they are part of something bigger and older than themselves—someone who is interested in transcribing manuscripts that are hundreds of years old may well be charmed, or at least intrigued, by the title amanuensis or scribe. Old Weather also uses this to great effect: the names of the levels reflect the content of the work volunteers are doing and may allow them to feel like they are part of the historical moments they are transcribing. Even intangible rewards like these make members of the crowd feel engaged and a sense of ownership with their project; archivists need to investigate the ways we can gamify our digital projects and make them more appealing to the crowd.
The case of the Ransom Center Fragments is slightly different, in that it grew out of one person’s curiosity and work. More than the other two case studies, too, this project required specialized skills—medieval handwriting is often more difficult to decipher than later scripts, and many, if not all, of these fragments must have been in Latin. I hypothesize that much of the appeal of this project to the community who worked on it was the rarity of the material—due to the ravages of time and history, simply less medieval material is extant than is from later periods. The discovery aspect of the Ransom Center Fragments was also appealing—there is a mystery to uncovering very old writing that may not have been seen in hundreds of years, and a thrill to deciphering and identifying it. Archivists can learn to seek out and engage with specific and specialized crowds who, in addition to having the necessary skills to do the work required, may be more enthusiastic about the task at hand than members of the general public.

Archivists can also learn, of course, from the Ransom Center Fragments that collaboration is essential to keeping crowdsourcing projects alive. Employees and team-leading volunteers must work together with the funding entity or entities to maintain projects over time, and there must be supports in place for when leaders feel unable to continue with their work. Archivists, along with other GLAM professionals, must plan for the human sustainability of their projects along with electronic and analogue sustainability. Alexandra Eveleigh argues that the role and labour of archivists can be undervalued when the crowd volunteers to do some of their work for them: “The bureaucratic nature of the authority wielded here is not relative to a particular sphere of interest or expertise, and hence does not of necessity have to be exercised by
professional archivists, and the reductive nature of the transcription task is easily dismissed as beneath the professional dignity of an archivist.”

However, characterizing the input of archivists in crowdsourcing projects as merely “bureaucratic” undersells the important work involved in constructing and maintaining these projects. The nature of these projects means that, in order for them to succeed, they need input from archivists, specifically postmodern archivists who do the kind of work Terry Cook describes:

…mediation by the archivist in setting standards, undertaking appraisal, targeting acquisitions, imposing orders of arrangement, creating logical descriptions, and encouraging certain types of preservation, use, and public programming is critically important in shaping [meaning relative to the context of the creation of the record].

Indeed, the assertions of Cook and of archivist Brien Brothman, who points out that “social communities create and destroy value,” mean that archivists must reach out to the communities surrounding them and their archives if they are to adequately and accurately describe their holdings. It is not enough, however, for archivists to merely approach communities and hope to preserve their heritage. In a 2011 chapter, Elizabeth Yakel describes the “First Great Opening” of the archives that took place in the 1960s as archives loosened restrictions on who could see their holdings and why. Yakel argues that 21st-century archivists forget that there was a time in the recent past when “access [to archives] was routinely denied due to such criteria as status and...
Historically, archivists have tended to guard their holdings too jealously, requiring letters of introduction or academic affiliation before allowing researchers to access their archives. As archival holdings became more diverse, reflecting the lives of more than just the most dominant group in society, so did the people seeking to learn from those archives. With the advent of social media technologies, Yakel posits a “Second Great Opening” of the archives, where archivists must learn to open themselves still more to the opinions and experiences of non-archivists. These changes can be disconcerting; as Yakel points out, “we are facing a change in the relationships between the records and the researchers that leaves out archivists.” Archivists need not be totally left out of this relationship, however. We must now contend with the ways we informally deny access to people whose histories and stories do not fit within traditional archival description as performed by archivists. Reaching out to a crowd—who are, after all, themselves users of archives—for help describing collections is one way to amend this.

Micah Erwin asserts that crowdsourcing projects, if they are to succeed, need stakeholders from an established institution. In addition to this, however, they need what Yakel calls “cognitive authority,” “the trustworthiness and reliability that people grant to texts, records, institutions, and people.” Institutions like archives have this authority in part because of the records they hold, and because of the ways those records are seen to be authentic. Although some of this authenticity comes from the institutions that sanction the official repositories that control their records, some of it also comes from the records themselves and the people who have taken care of them. As Yakel puts it, “archives maintain the authenticity of records through a chain of

---


187 Ibid., 77.

188 Ibid., 80.
custody and retain a certain moral authority by representing these records accurately." It is this authority that archivists are loath to give up, and it is this authority that can seem to be lacking in work performed by the crowd. Does an archives lose this authority if it shares it with members of the public? How much can the public contribute before the authority of the archives gets diluted? How can archivists walk the line between user-generated content and the “official” voice of the archives? As Yakel points out, “[h]ow best for archivists to be of, but not totally in, the community still needs to be worked out.” Archives and their holdings need to have cognitive authority if they are to be meaningful; the professional designations of archivists are one way that archival material and its descriptions can maintain this authority.

As discussed earlier, description is one of the most crucial tasks an archivist does. Opening up archival description to other users does not threaten the authority or expertise of the archivist; it merely recognizes the truth that the archivist is not an all-knowing objective authority. In addition to this, reaching out to members of the public to essentially create and enter item-level metadata for the made-digital material allows the made-digital to exist at the same level of discoverability as born-digital material. Material that has only ever existed in a digital environment has metadata attached to it (e.g. PREMIS events, format information) that can be repurposed as a description for the material. Scans or photographs of analogue documents have none of this information attached and so can be harder for researchers to find. As Lucinda Blaser points out, “[m]useums have started to become aware of the limitations of their knowledge: no employee can know everything there is to know about the institution’s collection or its subject

---

189 Ibid., 78.
190 Ibid., 88.
history.”192 This can be applied equally to archives and archivists: regardless of the training of any individual archivist, they cannot know everything about the archival material for which they are responsible. The online space of digital description allows room for other viewpoints to be heard, viewpoints which are just as valid in some ways as the more structured description written by the trained archivist. Blaser asks, “[i]s an official space for the official expert voice only, or can interpretation include the audiences and even grow as audiences in the spaces change?”193

The key to a collaborative description process for archival material is the multiplicity of voices that can contribute to users’ (and archivists’) understanding of the archival objects. Isto Huvila refers to this process as “decentralised curation:” “harnessing knowledgeable users of archival collections to contribute in the form of new and improved descriptions, translations, summaries, and relationships to other records.”194 Broadening the locus of authority in archival description in this way also broadens the extent of information that can be known and shared about each record. User participation in archival description can even shed light on records in ways that crowdsourcing organizers don’t expect: in Australia’s Your Archives project, as with Old Weather, the organizers were surprised at the ways volunteers contributed. “[T]he user community has been using Your Archives in ways we hadn’t considered: researchers have been writing transcripts, abstracts, and indexes of digitized sources, opening up the archives in a way that wasn’t possible before.”195 Allowing volunteers to contribute in a variety of ways lets

---


193 Ibid., 47-48.


archives benefit from the breadth of knowledge their user base has; limiting the contributions the crowd can make may also limit the engagement of users overall if volunteers feel they have contributed all they can, or that their specific knowledge is not relevant to the goals of the archives.

Terry Cook hypothesized that “postmodern descriptive systems would move away from the monolithic legacy of past archival theory, from ‘the old-fashioned “one-thing-one-entry” approach’ if they are intent on ‘satisfying researchers’ needs to understand the historical context of records, the activities that generated them, and the information they contain.’” Satisfying these needs is part of the mandate of archives and archivists, and using crowdsourcing to incorporate the knowledge of people beyond the archival profession is a way to respond to Cook’s argument. The projects described in this chapter transcend the “one-thing-one-entry” approach by acknowledging the limitations of archivists and working with their strengths to gather and share information that provides important context for archival material. Moving beyond a rigid vision of archives as the sole source of cognitive authority and sharing that authority with members of the public is crucial to making archives accessible and meaningful to future generations of researchers.

Case Study

As the projects discussed in this chapter have shown, it is possible to build a thriving crowdsourcing project, but doing so takes careful planning. For the hypothetical crowdsourcing project surrounding the manuscript fragments at the University of Manitoba, we can draw on the lessons learned from these projects.

The gamification strategies of *Transcribe Bentham* and *Old Weather* have been successful in retaining a crowd of contributors, though, as mentioned above, having leaderboards turn over more quickly would be likely to make the project more welcoming to newcomers. Ideally, our project would take inspiration from these two to allow members of the crowd to see how much they have contributed, and how their work stacks up against others’. In contrast to *Transcribe Bentham* and *Old Weather*, the texts being transcribed here are very short; an overall leaderboard along the lines of *Transcribe Bentham*, showing the total number of pages and boxes transcribed, would be less helpful here. Rather, we might count these contributions by fragment, or by word or line.

*Transcribe Bentham* is also inspirational in the ways it engaged with its audience and changed its user interface to suit their needs. Seeking and responding to user feedback is important when creating a crowdsourcing project, and the way *Transcribe Bentham* adapted to the needs of its users both improved the quality of the transcriptions and the happiness of the project’s contributors. The use of paid quality checkers in this project is also something I would want to include in the University of Manitoba fragments project—while this would incur some costs, it would be far less expensive than simply hiring someone to transcribe all of the material, and it would ensure that the transcriptions were accurate before disseminating them to the public.

One aspect of the *Transcribe Bentham* project that would not carry over to the University of Manitoba fragments project is the lack of any space for contributors to discuss the project. The team behind the *Old Weather* project noted that their users discussed the ships they were following on the project’s message boards, while Micah Erwin created a Facebook page where the contributors to the *Medieval Fragments Project* could discuss their and others’ work. While maintaining and moderating forums could prove to be additional labour for the people behind the
project, in addition to being another account for which contributors must sign up, using existing social media spaces, like Facebook, is a smart way to bring projects to the people contributing to them. While not all contributors may have Facebook accounts, it is likely that enough do to make using Facebook for this purpose an ideal solution.

Micah Erwin also used existing infrastructure to host the Medieval Fragments Project: by taking advantage of the free service provided by Flickr (though Erwin may have used a paid account for the project), he was able to keep costs for the project low. In addition to the lack of cost for the image and description hosting, using an external service to host the project meant that no one at the project was responsible for technological support: any issues with the website were Flickr’s problem, not Erwin’s. A problem with using external software is that creating a leaderboard for contributions would be more complex compared with Transcribe Bentham’s MediaWiki software; another is that, while the current Terms of Service for Flickr stipulate that Yahoo! (Flickr’s parent company) will only use your content “solely for the purpose for which such content was submitted or made available,” Terms of Service often change, and the rights to the images uploaded to a service like Flickr may also change over time.197

Finally, none of these tactics are worth anything if the project cannot attract a crowd. Based on Micah Erwin’s success gathering a crowd to work on the Ransom Center Fragments, I anticipate that the University of Manitoba fragments would not have a problem doing the same. Existing social media groups—listservs, Facebook groups, and Twitter—can be used to get the word out about the project. Because the crowd we are seeking is likely to be specialized in the field of palaeography or, at least, in medieval literature and textuality, making the project easy to

find and participate in will be more important than gathering a large number of contributors. By making participation in this project simple and painless, we can take advantage of people’s natural curiosity and interest in discovering new things about these fragments and the texts they contain.
Conclusion

Archivists arrange and describe material to make it accessible. In this core mission we are not alone: digital humanities and textual scholars have also, in their own ways, been working toward the same goal. Our work all fits together. Our projects complement each other and allow us to work together to provide the material this thesis is concerned with—digitized versions of medieval and early modern manuscript texts—with the best possible representation of itself. The *Rules for Archival Description* can and should learn a lot from these allied fields, especially as the Canadian Council of Archives works to update it, but there are lessons individual archivists and institutions can take as well.

Archivists already do many things that our digital humanities and textual studies colleagues can help us with: our material is often described in multiple locations at once, while digitized copies of it exist online in a separate location. Learning lessons from digital humanities scholars can help us to link these descriptions and images together in a way that makes them more discoverable and understandable to users of archives. For example, at the University of Manitoba Archives & Special Collections, archival fonds and collections are described in an instance of AtoM that complies with RAD. From that description, however, it is possible to link not only to an image of the item, but to a description of the item’s materiality that is encoded in TEI and may include crowdsourced information. Looking at the history of textual scholarship and the way it works to capture the physicality of texts can show us which information is important to include in these descriptions. By working together like this, we can provide researchers with not only a finding aid to the material in a collection, but an information package that describes many aspects of the items themselves.
Crowdsourcing projects not only help archivists describe material they may not have the resources or knowledge to, they can also attract a group of users that may not have engaged with the archives in the same way before. Indeed, reading a text closely the way crowdsource volunteers do is itself a form of using the archival material, ensuring that the material is read and engaged with at the same time it is being made more discoverable and useable. In addition to this, opening archival description up to non-archivists helps to foreground the subjective nature of archival work: the material that is archived was not archived for no reason, or by a vague institutional mandate. A person did that work, and crowdsourcing, by allowing others to help with that work, helps them to better understand its processes.

If archival material is not accessible, why is it being archived? Description is at the core of the archival mission because without description, archival material is lost. The material archivists choose to preserve is important and it is crucial that others be able to access it. Digitizing this material helps with the problem of physical access to archives, but users must be able to find what they are looking for. By working with and learning from our colleagues in digital humanities and textual studies, and by reaching out to members of the public, archivists can give the material we know is important the description it deserves.
Bibliography


http://idhmc.tamu.edu/arcgrant/.

“ARC | TypeWright.” Accessed February 6, 2015.
http://idhmc.tamu.edu/arcgrant/software/typewright/.


hftp://tapor.uvic.ca/~mholmes/image_markup/.


Horsman, Peter. “The Last Dance of the Phoenix or The De-Discovery of the Archival Fonds.” Archivaria 1, no. 54 (January 1, 2002).


Nesmith, Tom. “Archives from the Bottom Up: Social History and Archival Scholarship.” *Archivaria* 1, no. 14 (January 1, 1982).  

———. “Reopening Archives: Bringing New Contextualities into Archival Theory and Practice.” *Archivaria* 60 (September 25, 2006).  

doi:10.17723/aarc.65.1.rr48450509r0712u.


https://www.nedcc.org/preservation101/session-7/7digitization.


http://medievalfragments.wordpress.com/2013/04/05/navigating-the-digital-world/.

http://medievalfragments.wordpress.com/.


———. “The Ends of Editing.” *Digital Humanities Quarterly* 3, no. 3 (September 29, 2009).


Rydberg-Cox, Jeffrey A. “Digitizing Latin Incunabula: Challenges, Methods, and Possibilities.” *Digital Humanities Quarterly* 3, no. 1 (February 26, 2009).


——. “Humanities Computing as Digital Humanities.” *Digital Humanities Quarterly* 3, no. 3 (September 29, 2009).

——. “The Landscape of Digital Humanities.” *Digital Humanities Quarterly* 4, no. 1 (July 20, 2010).


Watt, David, and Elizabeth-Anne Johnson. “Medieval and Early-Modern Manuscript Fragments at the University of Manitoba,” n.d.


