The Journeys of Books:
Rare Books and Manuscripts Provenance Metadata in a Digital Age

by

Colette Leung

A thesis submitted in partial fulfillment of the requirements for the degrees of

Master of Arts in Humanities Computing
and
Master of Library and Information Studies

Humanities Computing / School of Library and Information Studies
University of Alberta

Abstract

This thesis seeks to examine the current state of the provenance metadata of rare books and manuscripts in digital special collections, and how that metadata can be enhanced using visualization tools. The multi-faceted nature of provenance is addressed, as well as the standards used to capture provenance metadata. Reasons for the development of these standards are identified and grounded in the historical development of both archives and special collections. Contemporary roles of provenance metadata in facilitating big data, interoperability, linked data, and data curation are also presented.

A two-part action research study was carried out. The first half of the study reviews the use of provenance metadata in sixty-four digital special collections, with focus on medieval and early modern rare books and manuscripts. Descriptive and digital factors of the collections are also considered in relation to the provenance metadata. Building on this research, visualization is tested as a means of addressing challenges in capturing provenance metadata and fulfilling contemporary uses of metadata for digital special collections. An environmental scan of eight visualization projects and case studies on five, open-source visualization tools were performed.

The conclusion finds that provenance metadata is in the best state it has ever been in both quality and extent of use, and that visualization has the potential to address issues in capturing provenance as metadata and enhance user experiences. Provenance metadata, however, still suffers from a lack of guidance, and would benefit from being recognized as several distinct fields instead of a single one. This thesis suggests what these fields might be, and reveals the potential to create a more robust visualization tool for provenance metadata. Directions for future research are also reviewed.

Acknowledgements

I wish to thank my supervisors, Ali Shiri and Harvey Quamen, for their guidance and support in preparing this thesis, as well as committee members John Considine and Dinesh Rathi. A special thanks is also owed to my friend and colleague, Glen Donlevy. Some of the early research for this thesis was inspired by our work together, and I will be forever grateful for that experience. I would also like to acknowledge the Government of Alberta Graduate Student Scholarship and the Queen Elizabeth II Graduate (Master's) Student Scholarship for the financial support received through those awards over the course of this thesis. Finally, I would also like to thank my work supervisors, colleagues, friends, and family who offered their support during my writing.

Table of Contents

List of Figures

List of Tables

INTRODUCTION

The aim of this thesis is to identify and evaluate current standards of provenance metadata, as used in Special Collections Libraries, and explore how current standards might be enhanced for humanities scholars using visualization tools in face of contemporary issues. Results from this study have implications for helping humanities scholars make sense of big data in the form of thousands of provenance metadata records, increasing interoperability between different institutional data sets, and creating linked provenance data. Results may also be helpful in designing Special Collection catalogues and potentially enhancing humanities research in both usability and findability.

Research Context

Metadata and Provenance

Metadata, or "data about data," has long served library settings in formal schemas for describing resources. Metadata can be applied "to any type of object, digital or non-digital" (NISO, 2004, p. 1). It is essentially structured information used to describe, locate, or make it easier to retrieve and manage an information item or source.

In the digital age, with the advent of the semantic web, big data, linked data, and federated searches, academic contexts are transforming. There is increased need for infrastructure to help manage and understand this data. Metadata is key in the process of organizing information, and supporting interoperability between systems. Many different metadata schemas have been created by various communities, and even within the same standard, metadata is variable, particularly as standards change over time. This is particularly true of special metadata fields, such as provenance.

Provenance is the history or chronology of ownership or location of an object, and is generally used to establish authenticity of that object. Provenance, explored more thoroughly in "Provenance Theory", has importance in diverse fields, including art, archival studies, rare books and manuscript studies, and computer science. It covers where and when an object has been, and who has interacted with it.

Although the concept of provenance dates back to the nineteenth century, it is an increasingly contemporary issue. "The serious study of private libraries, and of the lessons which can be learned from book ownership, is a growth industry and one which has gained much ground in the recent past" (Pearson, 1994, p. 2). Provenance has implications for data curation and preservation, not only for print objects in special collections, but also for digital items from a variety of fields. With the growing body of digital information, the value of provenance has re-emerged (Hartig, 2009), particularly in the form of metadata.

Capturing provenance as metadata draws from a long-standing and complicated history in both archival and special collection library practice, as a fundamental component of unique materials, and its consequential development across metadata standards. In the digital era, with over 300 digital libraries providing over 20,000 digitized medieval manuscripts (Menna & de Vos, 2014), scholars interact increasingly with digital objects, and less and less with physical rare book materials. Accurate and consistent information is essential for scholars working at a distance.

However, standards and use of provenance metadata across special collection libraries are highly inconsistent. Lundy (2008) observes that recording provenance is generally relegated to providing notes and added entries in bibliographic records, and in

some cases, previous ownership is not addressed at all. This is in contrast to the practice of archival institutions, as "significance of archival materials is heavily dependent on the context of their creation" (Hensen, 1989, p. 276). Online special library catalogues may not provide provenance metadata for multiple reasons, including difficulties in establishing provenance information, or lack of guidance in the cataloguing standards (Pearson, 1997). Many libraries, including government such as Library and Archives Canada and academic institutions like the Bruce Peel Special Collection at the University of Alberta, may also have provenance information in analogue format in old card catalogue systems, but due to cataloguing backlogs, this information is not yet online.

This creates an issue for many humanist scholars with focus on manuscript studies. If provenance metadata is not readily available, scholars must consistently reassess an object based on previous ownership. Catalogue records with provenance metadata have proven beneficial to scholars. Knowledge of provenance "lends authenticity and reliability to the materials held" (Buchanan, 2011, p. 349), but it also encourages scholars to "find new topics of research in special collections materials" (Lundy, 2008, p. 164). Other noted uses of provenance and related copy-specific information include added security to collections, determining future collection efforts, and issues of legality (see "Provenance Theory).

This leaves current special collections often facing the current scenarios:

- Provenance information, if known, has been applied in multiple formats and standards, sometimes without consistency (Lundy, 2008).
- The depth of which to display information is unsure (Pearson, 1997).

- Provenance information may be stored in backlogs, and considering current challenges, has no incentive to be prioritized (Buchannan, 2011).
- There is a growing amount of provenance metadata, which users need to make sense of, interact with productively in the digital medium (Elings & Waibel, 2007).

Examining provenance metadata is important not only for providing users with information clearly and quickly for their research needs, but also because of shifts in technology. The advent of the semantic web has changed the way scholars approach metadata, and has contributed to a systemic shift in how individual scholars browse and search for information. Collaborative and large-scale projects have emerged in light of new technology. Metadata, in the digital era, is part of the "bigger picture" that allows scholars to understand large amounts of data, and to utilize linked data. New questions can be asked of provenance metadata that could not before, such as tracking trends in the historical exchange of books, or rebuilding historical collections (McQuillen, 2013).

Current use of provenance metadata may also have impact on data curation, or the preservation of data or information for long-term re-use. The authenticity and history of digital objects will be important, especially if the original is no longer accessible if a library chooses to restrict use of the object for preservation purposes.

Metadata Visualization

Metadata, especially once standardized, can be used to build rich visualizations. Many digital collections rich in metadata have been released or are forthcoming. Libraries offer retrieval interfaces; however, many are looking towards developing more meaningful representations, especially with the development of Next Generation

Catalogues. In a world where physical objects are increasingly digitized, libraries must also engage with available technological tools, in order to meet audience needs, and best represent materials. These are some of the goals of Next Generation catalogues. As a consequence, libraries tend towards taking advantage of rich digital tools to help users understand, browse, search, and organize a collection.

Visualizations are useful tools when dealing with large data sets. They allow scholars to understand vast amounts of information quickly (Araya, 2003), and to make connections between data that would have been impossible before. In other words, a display can influence users to think in a particular way about the collection. Visualization and provenance techniques have rarely been combined before, but considering the inherent visual metaphors of provenance in time, place, networks, such a combination in digital maps and timelines may be fruitful in furthering understanding of book history.

Some of ways visualizing provenance metadata may prove worthwhile include:

- The standardization and cleaning of data, which must be done before it can be used in visualization (Van Verchum & Pugin, 2004). In other words, using visualization tools may help provide some guidelines in standardizing provenance metadata.
- The ability to make sense of large data sets (such as 20,000 digitized manuscripts). As more data is created, humanities disciplines are looking less towards saving or deleting data, but rather finding ways to use large amounts of data (Chen, Mao, Zhang & Leung, 2014). Visualization could organize “big” data in the humanities.

- The ability to determine and create links between data in ways not considered before (Williford & Henry, 2012a). For example, a visualization may allow one to literally see a connection between two data sets based on cross-sections in time and place.
- The increased promotion of the interoperability of data sets between different collections, as provenance in multiple standards could be harvested and standardized (Mazza, 2009). Research is no longer the realm of only the lone scholar, but instead focuses on collaboration of data, particularly between collections and different institutions (see Williford & Henry, 2012b).
- Visualization may increase usability, access, and information retrieval of both physical and digitized materials (Shiri & Molberg, 2005) in Special Collections.

One of the challenges libraries face when look towards creating Next Generation catalogues and employing digital tools is using existing metadata that is either poor or inconsistent. Cleaning metadata is an important step of any visualization. This thesis will establish the extent standardization might be undertaken with current digital tools.

Research Questions

The multiple standards of provenance metadata and the increased amount of digitized data have led to a number of timely questions currently in need of answering, especially as libraries move forward in the digital age and seek to make this data accessible to users on new levels. These questions outline the conducted study.

1. What standards exist in provenance metadata? Why did so many come to exist and how do they differ?

2. How are Special Collections libraries currently using these standards? Are current standards meeting current needs?
3. How might provenance metadata be displayed most productively to users on digital interfaces using visualizations?
4. How can visualizations help resolve current issues surrounding provenance metadata? Do they meet user needs?

Thesis Outline

In order to answer the outlined research questions, this thesis will begin by examining a brief history of provenance. This includes further exploration of the importance of provenance, and the history and communities which inform our current understanding of provenance theory. "Provenance as Metadata" discusses current provenance metadata standards available to institutions, and examines modern understandings of metadata and its contemporary digital uses.

"Methodology" presents the overall methodology and outline for a two-part study that builds on the history of provenance to understand the current state of provenance metadata and how provenance might be effectively visualized. This thesis is qualitative in nature, and uses multiple methods in order to approach provenance with a holistic design. The overarching methodology of the two-part study is action research based. More specifically, the first part of the study uses comparative and evaluative methods on a sample of sixty-four digital special collections, to examine the current use of provenance metadata. The results of this first half of the study are discussed in "Provenance Metadata in Digital Special Collections". Analysis is also provided through historical investigation in the literature review.

The second part of the study provides a brief environmental scan of existing visualization projects, and short case studies of available visualization tools. These tools are used in a pilot test with sample provenance data, in order to assess their potential to resolve issues in capturing and displaying provenance metadata. Assessment of tools is based on usability approaches, modeling, and systems analysis. Results from the second half of the study are discussed in "Visualizing Provenance Metadata". The conclusion summarizes major findings of this thesis, including recommendations for special collections and future lines of study.

The scope of the project will include manuscripts and rare book collections with at least a partially digitized collection, and with an emphasis on works in the medieval and early modern period (before the 1700s) in Europe. As such, provenance theory in non-Western countries falls beyond the scope of this project.[1] Rare books, defined as those with only a few known copies or that are unique in some capacity, and manuscripts have been selected to create a focused study. Born digital items will therefore be excluded.

Results of this thesis may have implications for both scanned and born digital objects, and may be helpful in designing Special Collection library catalogues. It is especially timely considering the introduction of RDA cataloguing standards into most academic and public libraries, the transition towards Next Generation Catalogues (NGC), and the increased amount of metadata being generated (see "Provenance as Metadata"). This study crosses over with library studies, digital humanities, and book history.

---

[1] A project that seeks to encompass non-Western provenance should pay attention to alternate understandings of mapping, and different provenance markings. For example, in many East Asian manuscripts, the use of seals and allowances for those kinds of marks should be emphasized.

CHAPTER TWO: PROVENANCE THEORY

This chapter seeks to provide a basic history of provenance theory in order to establish a basis for understanding modern methods of capturing provenance metadata. Tracing the history of provenance is not a direct task. Multiple disciplines have used the term "provenance" in slightly different ways, and the use of provenance is constantly evolving. Examining how the concept of provenance has evolved over time, to the point it has reached today, helps us understand interdisciplinary influences on the theory, and the directions provenance may take in the digital age.

The word "provenance" itself derives from the French term 'provenir,' meaning "origin", "cause", or "to come from" (provenir, n.d.). It was first used in English with this meaning in the late 18th century. It was not until the late 19th century, however, that the word provenance was used with regularity in many disciplines, and a more nuanced meaning developed.

Although the roots of the word provenance come from the idea of "origin," the meaning of provenance expanded to include multiple steps of ownership and associated information (e.g. date, place, etc.). This began during the 20th century, when provenance was used with professional meaning as the chain of ownership, often as a documented record used to prove authenticity or quality of a special object (Cook, 2013; IFAR, n.d.). It is likely this term carried over from French archival use of the word "provenance", which was popular at the same time. By the mid-1900s, provenance was regularly used in this context in art and archaeology (IFAR, n.d.), reflecting the establishment of provenance theories within these disciplines.

The established use of provenance as a formal aspect of an object by the 20th century helped launch the theory of provenance in a multitude of disciplines, including not only art, and archaeology, but science, archives, manuscripts and rare books, and computing (Davis, 2014). Each of these disciplines uses provenance in slightly different yet related ways. Surveying the meaning of provenance across multiple disciplines informs the elements at the core of provenance theory (see *Table 1*).

Table 1

The Definitions of Provenance According to Discipline

Discipline	Definition of Provenance
Art and Antiquities	Methodology for reconstructing ownership (Davis, 2014; IFAR, n.d.; Stein, n.d.).
Archaeology	Chain of Ownership; Placement of objects within excavations (also known as provenience) (IFAR, n.d.).
Palaeontology	Records of ownership of fossils; Where a fossil was found (American Museum of Natural History, n.d.).
Geology	The layer of ground where the rock in question was found (Ingersoll, 2014).
Botany	Place of origin of a seed; A seed's genetic history (Wickens, 2004).
Wine	Where and how wine was stored, and for how long (Christie's, 2009).
Law	Related to "chain of custody"; especially important for issues of copyright and plagiarism (Carless, 2015).
Science	Knowledge of steps taken in a scientific experiment, allowing repeatability of data (Davis, 2014; Groth, 2007).
Computer Science	Data Provenance including workflow, origins, and creation of data. The standard associated with provenance in computer science is PROV, designed with the semantic web in mind (Hartig, 2009; Niu, 2013; W3C, 2013;).
Archival Science	Founding principle for the management and organization of archival records, which suggests that the order records come to an archive in should be preserved, and that records should be organized according to ownership (as opposed to subject, etc.) (Niu, 2013).
Book History / Library Science (Special Collections)	Ownership of books, including the evidence left by owners on those books, and other contextual evidence such as where and when a book has been (Pearson, 1994).

Table 1

It becomes clear from examining the definitions in *Table 1* that the core elements of provenance include not only ownership, but contextual details such as where, and when an object has traveled. While detailed examinations of provenance in most of these fields extends beyond the scope of this thesis, it is important to recognize that as more data from all disciplines becomes digitized and accessible through the web, there will be increased interaction between the many interpretations of provenance (Niu, 2013).

For the purposes of this thesis, provenance refers to "both the origin of an item, and the history of ownership of that work" (Greenberg, 2005, p. 17), as applied in the contemporary field of book history. In this context, provenance is generally expressed as a chronology of owners, custody, or location of an historical book. It includes not only names, but other contextual information of interest to book history (such as time and place) to allow further understandings of people's interactions with books over time.

Importance of Provenance

In book history, provenance marks are potentially the "most valuable pieces of documentation relating to a collection as a whole" (Buchanan, 2011, p. 349). Provenance is important across disciplines for a number of reasons. The most cited is probably authenticity (Buchanan, 2011; IFAR, n.d.; Stein, n.d.) – knowledge of provenance helps guarantee that an item has not been forged and is an accurate representation of history. In digital mediums,[2] the importance of provenance is amplified, as scholars may work at a distance, and need to be sure the information and materials they are accessing digitally are sound. By extension, the provenance of a rare book may increase its monetary value (IFAR, n.d.; Lundy, 2008; McQuillen, 2013) – if an important person owned the book,

[2] This thesis favours the use of "mediums" as the plural of "media," over the Latin plural "media."

collectors may find it more valuable. The *ABC for Book Collectors* (2004) makes note that provenance is often tied to the quality of a book (Carter & Barker, 2004).

More recently, provenance has become important in regards to rare artifacts due to issues of legality and ownership (IFAR, n.d.; Stein, n.d.). An accurate provenance proves that works were legally acquired (American Museum of Natural History, n.d.), or if works were stolen, provenance allows an institution to return them to their rightful owners (Stein, n.d.). This particular aspect of provenance has become important in light of cultural heritage, and cases such as stolen art under Nazi Germany (Stein, n.d.).

On a more macro level, provenance may also provide information about historical aspects surrounding the book, including the owners of books, how people historically interacted with books, and aspects of the larger discipline of book history (IFAR, n.d.). For example, provenance may provide a record of social, economic, and political networks, the influence of books over these, and how those networks changed over time (Stein, n.d.). By extension, provenance can also help show different cultural attitudes towards ownership (for example, Asian marks of ownership differ from European ones – where objects of European origin often have gaps missing in provenance, previous owners, even if a relatively unknown figure, are important and recorded through use of unique seals).

Provenance may also provide information on which books were influential to important or well-known historical figures in various disciplines (Lundy, 2008), although "neither the books nor the former owners have to be rare or famous to make worthwhile the effort of cataloguers to record provenance information and make it searchable in bibliographic records" (Lundy, 2008, p. 164).

Provenance therefore becomes especially meaningful in highlighting relationships between objects, people, places, time periods, and other objects. "[R]ich paths of discovery are found and validated in interconnected and related connections" (Stein, n.d., para. 4) uncovered by provenance. By extension, provenance theory also has specific impact on libraries and the study of library history. If provenance information provides social context to a book, it also has the potential to highlight links between collectors and larger institutions (Oldfield, 1994). The study of provenance provides information about the "fashions and tastes" that historically influenced libraries and collections (Oldfield, 1994). In further practical terms, provenance of rare book collections may help guide future acquisition on the part of special collection libraries to fill needed areas (Buchanan, 2011), and may influence what information or items researcher request of their institutions (Buchanan, 2011).

Provenance theory also lends itself to the research conducted by scholars by creating new areas of interest (Lundy, 2008). For example, Overmier & Doak (1996) note that Fermat's Last Theorem, recorded in the margin of a book he owned,[3] suggested that a certain mathematical problem should be easy to solve, although did not provide a solution. This singular note inspired mathematicians to study the problem for centuries (Overmier & Doak, 1996).

Provenance theory can also provide many rich aspects of study when large amounts of data are examined concurrently. For example, provenance has recently been used to reconstruct the holdings of former library collections, such as McQuillen's (2013) reconstruction of the Scheyern Abbey library collection from Bavaria. This

[3] For the purpose of this thesis, copy-specific information is considered closely related to provenance evidence, as both provide evidence of how a book was used or of ownership.

reconstruction allowed a new examination of the place of the book in fifteenth-century Bavarian culture and society (McQuillen, 2013). Similarly, Rueda-Ramìrez and Ruiz (2009) used provenance records from the Historic Collection of the Library of the University of Barcelona to reconstruct the collection of former convent libraries. This research led to further understanding of the role of those libraries in the cultural history of Spain, and the role of convents as public libraries (Rueda-Ramìrez & Ruiz, 2009).

This kind of research means that provenance has taken on further importance for libraries and archives in the digital medium, because of the large amounts of data becoming available. It is believed that digitization has reached "a point where any English-language book printed before 1700 is available as a full-text digital facsimile, and soon that could extend to 1800" (Pearson, 2005, p. 4). These endeavors are brought about not only by libraries and other institutes digitizing their works, but also by outside interest through projects like Google Books (Pearson, 2005).

Providing provenance information across multiple institutions is of interest for sharing data, solidifying digital collections, and creating improved searches and research.

Fundamental Elements of Provenance

Although provenance means many different things to many disciplines, in book history, provenance tells us the journey of a book – who owned it during what time and where, and how the item changed hands or was read at that same time and place. From this understanding, the fundamental elements of provenance can be identified, and seen repeated across certain disciplines (such as archaeology, art, the study of wine, and law).

Three consistent axes are identifiable in provenance theory for every object:

- owner(s),
- time(s),
- and place(s).

But further to that, and perhaps most importantly are the relationships between these three axes (to other owners, other books, etc). By its very definition, provenance conveys information about the connections books have to their creators, owners, institutions, and other books that may have interacted with them through their journey to whatever special collection they come to reside in.

These three axes of provenance – time, space, and ownership – and the rich relationships between them, help layer rare book provenance theory on many complex levels, which in turn affect how different communities have chosen to record provenance information (see "Provenance as Metadata"). Further, the complex relationships between these three axes have become an increased focus of both archival theory, and special collections mandates, leading to the need for methods of highlighting and explaining those relationships.

The following section of this thesis examines how provenance of a book is assessed, and what challenges may present themselves in that assessment. Afterwards, a history of provenance theory through to the present day is presented, so that modern assessments of provenance might be better understood.

Assessing Provenance

Before analysis can commence on rare book items using provenance theory, scholars must first begin with provenance research. Such research involves the identification "of the who, what, where, and when" (McQuillen, 2013, p. 1) of a book, in

other words, trying to establish the three axes discussed above. Establishing provenance of an item is not easy, as becoming "literate" in provenance involves reviewing any one item for (Pearson, 1994): inscriptions and manuscript additions (marginalia), bookplate, book labels, book stamps, bindings, and sale or library catalogues.

Any single one of these aspects, as well as any combination of them, must be assessed in order to fully determine provenance. Even traces from anonymous people are of importance (Scheibe, 2010). David Pearson's comprehensive guide, *Provenance Research in Book History: A Handbook* (1994), remains one of the most important and authoritative sources available to help scholars with provenance research. These aspects are described in much greater detail in Pearson's work, and should be consulted by researchers beginning provenance research on a specific book or manuscript.

Provenance research may therefore involve consulting resources beyond the book itself, such as documents and archival resources. These might include gallery correspondence, stock inventories, estate records or histories, sale ledgers, photographic images, newspaper articles, as well as exhibition and bibliographic materials (Stein, n.d.). More recently, oral histories have also been used to help reconstruct provenance (Stein, n.d.), particularly for cultures which use oral histories to capture their history. Multiple institutions have created many repositories in order to help provide some of this information to researchers (Stein, n.d.). These are sometimes referred to as "provenance databases", and include armorial and ownership databases. The number of these repositories has grown, especially online in recent years. Examples include the English Short Title Catalogue (ESTC), the British Library's Database of Bookbindings, and the Consortium of European Research Libraries (CERL) Databases.

Recent scholarship has suggested that provenance research should be conducted backwards in time – researchers should begin with the known collection the book is part of, and trace each step back methodically (van Heel, 2011). This helps ensure a more comprehensive provenance. In general, scholars agree that provenance should be included in collection catalogues (van Heel, 2011) although the level of detail institutions should represent is a matter of some debate (see "Provenance as Metadata").

A number of potential challenges exist in provenance research. Reconstructing the past history of an object may be difficult – there might be multiple accounts of where an object has been. Alternately, records are often incomplete, leaving long period of time unaccounted for in the history of an object (Pearson, 1994). Although educated guesses might be made in some instances for some objects, it is sometimes impossible to reconstruct where the book has been during a certain period of time, or who owned it. Other times, extensive research is required in order to establish the missing information.

Because provenance is a concept with multiple layers, it is also possible for some information to be known, but not all. For example, an owner might be known, but it might be difficult to find out where or when a book was owned, if it can be found at all – owners may have moved during their lifetimes, or lived in multiple residences. Books are further complicated as they may have been separated or recombined with other books, providing them with multiple provenances (Pearson, 1994). For example, old manuscript pieces are often used in the bindings of later books, such as the case of *Commentarius juridico-historico-politicus de comitiis Romano-Germanici Imperii* at the University of Alberta. The outermost layer of this treatise on law's binding is a reused liturgical manuscript (Brown & Considine, 2010).

In addition to the history of an item itself, the history of provenance theory in archives and special collections informs how provenance is captured.

Provenance History in Archival Theory

Provenance has a long-standing and complicated history in both archival and special collection library practice, as a fundamental component of unique materials. Both the nature of archives and special collections has affected the development and understanding of provenance, and both disciplines have ultimately affected how provenance has been generated as metadata (see "Provenance as Metadata").

Prior to the eighteenth and nineteenth century, the demarcations between libraries and archives were much more blurred. However, as archives developed their own identity over the past two hundred years, they did so in connection with theories of provenance. These theories are often well-structured, but have changed over time and place.

Although full explorations of provenance in archival theory have been described elsewhere (see Cook, 2013; Douglas, 2010), a short summary is provided here so the concept can be better understood in special collection libraries. In the digital realm, the distinctions between libraries and archives begin to blur again, especially as users search for information through more general interfaces, such as online search engines. In order to understand the current climate of provenance in special collection libraries, an understanding of the history of provenance in archival theory is also necessary.

Most European countries can trace their own separate path of archival development (Duchein, 1992). Most early archives, however, were interchangeable with early libraries, and served as collections of documents. Provenance was not initially considered. Many early archives were religious by nature.

By the end of the eighteenth century, however, the archival field was fundamentally changed, due to Napoleon I and the French Revolution (Duchein, 1992). At the time, many legal structures in Europe were undergoing significant change, and accordingly archives began to change.

Due to the large change in political and legal structures at the time, many of the holdings in archival repositories became outdated as they contained information about old forms of government, monasteries, etc. Archives became largely historical in nature because of this (Duchein, 1992; Posner, 1940). The other major change instigated by the French Revolution was the uniting of archives as national institutions, which helped develop national identities of many countries (Duchein, 1992), although each archive maintained their own methods and standards of describing materials.

One of the outcomes of changes in archives over the course of the eighteenth century was the employment of scholars in archives, to arrange and catalogue materials (Posner, 1940). Archives developed "to a certain degree along the line of manuscript collection libraries" (Posner, 1940, p. 166-167) as archivists generally trained in libraries.

Towards the nineteenth century, archival science began being offered as a subject in many European schools (Duchein, 1992). These schools were different from modern archival schools. Subjects taught to prospective archivists included history, paleography, philology, diplomatics, sigillography, and medieval heraldry (Duchein, 1992) – skills more associated with modern book historians, rather than modern archivists. The roots of archives remained tied to special collections even as archives began to emerge as their own institution.

The shift towards modern archives, and ultimately provenance theory, began in the nineteenth century. In 1808, papers from the Public Works Division of Prefectures in France began to be transferred with regularity to archives as ordered by Napoleon I (Duchein, 1992). This process began to spread across Europe, moving archives out of historical spheres, and soon archivists were dealing with enough information that issues of appraisal, user access, arrangement, and description began to emerge (Duchein, 1992). Accordingly, by the mid-nineteenth century, archival science began to be taught independently of library subjects, at least in France (Duchein, 1992).

Archival arrangement became a hot topic in Europe (Duchein, 1992). Many archives followed systems likely borrowed from libraries (Sweeney, 2008), where records were arranged in ways to help support the research of scholars (Douglas, 2010). This meant that records were usually classified according to their subject, regardless of provenance (Posner, 1940). Classification, in general, was a popular topic during the age of Enlightenment due to expanding theories of evolution, social sciences, etc. (Sweeney, 2008), which may have contributed to the zeal of archival classification.

Early on, however, these systems were criticized for destroying the context of record creation (Sweeney, 2008) – an aspect that was increasingly important as an access point for users. This led to the development of the concept of *respect des fonds*, first properly defined in 1841 by the French Minister of Cultural Affairs, Natalis de Wailly (Bartlett, 1992; Duchein, 1992; Sweeney, 2008).

The term was coined in the *Instuctions pour la mise en ordre de la classement des archives départementales et communales* (Douglas, 2010), and it stated that "all documents which come from a body, an establishment, a family, or an individual form a

fonds and must be kept together" (de Wailly, 1841, as cited in Douglas, 2010, p. 25). De Wailly's *respect des fonds* suggested that records should be kept according to provenance in a *fonds*, and the order which records were received had to be respected. The time, place, and creators of records were all very important. A *fonds* cannot exist without this information (Cook, 1993). Already, the axes of provenance as reviewed above were becoming central to archival theory, and the use of the term began to emerge as recorded in the Oxford English Dictionary.

It is generally believed that *respect des fonds* was proposed to prevent damage to records, and to reduce the need for supervision of workers when putting records away (Douglas, 2010), but the concept served well to address criticisms of archival classification. Provenance became the foundation for archival arrangement (Duchein, 1992), and archives began to distance themselves from libraries, becoming more registry-like in nature (Posner, 1940). It is worth noting that by and large private and personal records were not initially part of archival traditions. Those materials were kept instead in libraries and special institutions, with exception of Canadian archives (Cook, 2013).

The status of provenance as a foundational principle in archives was solidified through a number of archival manuals, the most famous of which was the Dutch Manual, the *Manual for Arrangement and Descriptions of Archives*, by the authors Samuel Muller, Johan Feith, and Robert Fruin (Cook 1997; Duchein, 1992). The Dutch Manual, published in 1898, was "widely recognized as having gathered together and elaborated on ideas about the arrangement and description of archives that had been circulating in various European countries over the previous century" (Douglas, 2010, p. 27), the most significant of which was provenance. As provenance emerged as a foundational theory in

archival practice, the word also began to surface in the English language, and was concurrently used in other disciplines. Both provenance and *respect des fonds* remain the foundations of archival science (Duchein, 1992), and are used for arrangement, retrieval, acquisition of new records, appraisal of records, proof of authenticity, and description of records (Sweeney, 2008).

The Dutch Manual became one of the most influential works in archival theory. Provenance was further reinforced in other famous manuals, such as those by Eugenio Casanova, Sir Hilary Jenkinson, Theodore R. Schellenberg, and Elio Lodolini (Douglas, 2010; Duchein, 1992). Over the nineteenth century, provenance became a central concept in archival science across the world (Cook, 1997; Douglas, 2010; Sweeney, 2008). Even so, understandings of provenance have continued to change depending on time and place: "[w]hile international standards exist for the cataloguing of books in libraries, there is nothing […] even remotely approaching a common doctrine for the inventorying of documents in archives" (Duchein, 1992, p. 20)

The introduction of provenance as a foundation in archival science is perceived to help pinpoint the beginnings of delineation between archives and libraries, a gap that remains in a somewhat tenuous manner to this day. It is worth noting, however, that although many archivists maintain that libraries and archives are quite different, and have been so "since the beginning" (Duchein, 1992, p. 21), a more accurate history acknowledges that libraries and archives have a shared history, and maintained a complex relationship. In the 18th century, when archives began to be viewed as more historical institutions, some European libraries began to keep archives (Duchein, 1992). In fact,

many special collections libraries hold manuscripts - some of which are archival documents that could also be easily held in archives (Duchein, 1992).

Outside of special collections, however, archives were never controlled institutionally by libraries in Europe (Duchein, 1992); instead, different countries maintained different ways of running archival institutions, some of which were closely tied to libraries, and some of which were not (Posner, 1940). Consequently, archival practices influenced special collections, rather than the other way around.

Modern Interpretations of Provenance in Archives

Although provenance has served as a foundational principle in archival science, it has also changed meaning several times throughout the development of modern archives (Cook, 2013; Douglas, 2010). It has served as an organizing principle, an intellectual construct, and as sociohistorical context to the records in question (Douglas, 2010).

By the second half of the twentieth century, provenance began being questioned (Bearman & Lytle, 1985; Douglas, 2010; Posner, 1940). These critics of provenance argued that arranging records according to provenance does not allow for the changing nature of institutions (Bartlett, 1992). Records are not often subject to one-way hierarchies between organizations and records (Bearman & Lytle, 1985; Cook, 2013). Technology has affected provenance in archival mediums, encouraging more dynamic relationships between records, creators, and functions, including many-to-one, one-to-many, and many-to-many relationships (Cook, 1997). In other words, a single record could have many contributors, many functions, and relationships to many other records.

In response to these changes, some countries have re-assessed their use of provenance in archives. Many European studies have been conducted, reaffirming the

usefulness of provenance, although its interpretation is more liberal in modern times, especially for electronic records (Cook, 1997). In Canada, a "rediscovery" of provenance occurred in the second half of the twentieth century (Cook, 1997). National descriptive standards were created to capture the context archives were created in (including personal archival documents), and new archival studies programs were founded (Cook, 1997; Douglas, 2010). Around the same time, Australians similarly "rediscovered provenance" and adopted the Scott's Series System, which emphasized the multiple relationships of records to creators and other records (Cook, 1997; Douglas, 2010; Sweeney, 2008).

The development of archives proceeded differently in the United States than in other areas of the world, but its interpretation of provenance is worth summarizing. Manuscript collections, rooted in librarianship, were the first to adopt archival methods in the United States in the early twentieth century, and did so for decades (Hirsch, 2010), in spite of the fact that the Dutch Manual suggested manuscripts should be separate from archives as they were not strictly records (Hirsch, 2010). Provenance provided a way for manuscript libraries to evaluate and understand the historical context surrounding materials (Hirsch, 2010). In 1965, *The Management of Archives* manual was published, officially merging manuscript tradition and archives, but the merge was slow to occur (Hirsch, 2010). As in Europe, American archival institutions developed their own rules for many years (Hirsch, 2010). The profession still deals with tension from these different traditions today.

More recently, provenance theory in archival practice has undergone a further shift towards providing social and cultural contexts to records. This shift occurred due to increased academic interest in the narrative nature of records, and their history (Douglas,

2010). Scholarship became interested not only in the records themselves, but who created them, the community records contributed to, and when and where these interactions occurred (Douglas, 2010). The relationships between different aspects of provenance, became emphasized, bringing to light the importance of the axes described above.

Such interpretations of provenance have large implications for archival science, as they suggest that archives are not as "neutral" as the profession once believed (Cook, 2013). The methods used to acquire documents, describe them, and preserve them mean that provenance shapes concepts like memory, identity, and community. Provenance is the ground of complex interactions between libraries, archives, records, and users, and this rings true for provenance of special collections as well.

Most archival guides suggest that descriptions in *finding aids* (where most descriptions of provenance are captured) should include creator history, records history (how the material was managed over time), and custodial history (Douglas, 2010), although there seems to be lack of guidance in the consistent application of the field, or distinguishing between them (Bearman & Lytle, 1985). There has also been suggestion of a secondary provenance in some circles, as described by Lori Podolski Nordland. Secondary provenance would provide information on how the meaning and importance of records changes as their interpretation changes over time (Douglas, 2010).

More recently, there have been calls for expanded descriptions of provenance, due to new understandings of the concept. Although capturing provenance information is multi-layered and at times limited, it is agreed that the information should be captured to the best possible degree, in order to improve user access to records (Bearman & Lytle,

1985). There is a move towards the use of multiple metadata fields for provenance, as opposed to more traditional descriptive practices (Douglas, 2010).[4]

The meanings of provenance in archival theory continue to evolve (Douglas, 2010), especially as digital mediums allow for flexibility in recording and presenting provenance information (Sweeney, 2008). Archivists have called for involving communities in capturing provenance information (Cook, 2013). There has also been a push towards more conscious understandings of how users interact with archival systems, especially digitally, and how that may possibly affect provenance information (Cook, 1997). Part of this push would be for allowing many-to-many relationships to be displayed (Cook, 1997).

Suggestions that archival institutions have proposed to help meet these goals include creating links between multiple institutions (Cook, 1997), allowing for patterns to be identified in records, and management of big data (Cook, 1997). Multiple suggestions for the creation of a provenance database have been provided (Cook, 1993). Archives are also encouraged to "concentrate on mapping relationships existing within the nature of each and between the two. It is this very mapping which, bit by bit, part by part, reveals the whole" (Eastood, 1990, as cited in Cook, 1993, p33). Highlighting the relationships created by provenance is key in moving forward with the concept in archival theory.

Sweeney (2008) reflects that, "the term 'provenance' [… is] like a railroad train that picks up and discharges passengers at stations as it rumbles along its circumscribed path through the countryside" (p. 207). We have a continued fascination with archives

[4] It is worth noting that as of 2005, the International Council on Archives (ICA) Descriptive Standards Committee allows the combination of metadata elements, but multiple provenances have not been allowed for yet (Hurley, 2005).

(Derrida, 1995), and therefore provenance. The expansion of provenance theory becomes particularly necessary as archives and special collections begin to interact more in digital environments, and the use of metadata in digital tools opens new possibilities for highlighting the relationships seen as key in provenance theory.

Provenance History in Special Collection Libraries

The historical development of libraries has played a large role in shaping provenance theories in the discipline. Although libraries and archives developed in conjunction for centuries, libraries did not define provenance, or use it with regularity, until the twentieth century. Even then, provenance never took on the fundamental role it did in archival institutions. This is further complicated by the fact that many different kinds of libraries exist, and not all seek to capture provenance information.

Early archives and libraries both developed in tandem, usually as interchangeable institutions associated with power and authority (Hedstrom & King, 2003). This relationship continued as book culture grew over several centuries. By the end of the fifteenth century, the roots of provenance were being sown as private and institutional libraries were increasingly important (Summit, 2008). Many of significant libraries of the era would serve as the foundations of modern day special collections. These include the Parker Library in Cambridge, established in 1575; the Cotton Library in London, established in 1588; and the Bodleian Library in Oxford, established in 1602.

Early private libraries created some of the first rules for organizing medieval texts (Summit, 2008). According to Summit (2008), these rules

> still survive in the catalogues, classificatory schemes, and notation systems through which Renaissance collectors, readers, and librarians arranged and left

> their mark – sometimes liberally- on the medieval textual materials that they collected. (p. 2)

Many of these marks and protocols have large impact on understanding the provenance of medieval manuscripts today. During the seventeenth century, private libraries also began to focus on gathering unique items.

Over the course of the eighteenth century, libraries and archives continued to grow in numbers and scope (Hedstrom & King, 2003). Private collections were often given to universities or government entities (Hedstrom & King, 2003), supplemented by archival materials. By 1800, twenty national libraries existed across Europe (Hedstrom & King, 2003). For example, the British Library was founded in 1753 (Hedstrom & King, 2003). Overseas, the Library of Congress in the United States was established in 1800. The appearance of National Libraries coincided with the increased role of archives in the same nationalistic capacity. Libraries were experimenting with identity, and crossed over freely with archives.

In the eighteenth century, libraries also became involved in public book-lending (Summit, 2008), although this was not consistent across all institutions. Subscription libraries developed towards the nineteenth century, although rare book libraries tended to remain private (Summit, 2008). Around this same time, after the French Revolution, libraries and archives began to struggle with separate identities. Vast collections of archival materials, books, and manuscripts needed organization and structure after the revolution (Hedstrom & King, 2003), and tensions between the two institutions began to solidify. Libraries and archives separated, and the latter adopting the concept of provenance as both a method of organization and a foundational principle.

Libraries began to concurrently develop their own organizational methods, through cataloguing and classification systems. Major systems, still in use today, were created in the nineteenth century. It was a time of considerable development for libraries. Significant systems include Sir Anthony Pazzini's 91 rules for cataloguing, the Jewett cataloguing system (1853), and the Cutter Expansive Classification and the Dewey Decimal System (1876).

Library classification and cataloguing systems continued to develop in the twentieth century. The Library of Congress Classification was developed in 1901. When computers were introduced into libraries in the 1950s (Hestrom & King, 2003), new systems of metadata were needed again. MAchine-Readable Cataloguing (MARC) was subsequently developed in the 1960s, and the Anglo-American Cataloguing Rules (AACR)[5] in 1967. By 1971, these standards were used not only in the United States, but in many English-speaking countries around the world (Hedstrom & King, 2003). None of these systems were based upon provenance theory, as archives were, but over time, may would come to address provenance in some peripheral way.

The development of separate organizational methods in the nineteenth century helped establish libraries and archives as individual institutions. Separate institutions helped provide perceived authenticity and legitimacy to their collections. Many countries established both a national archive and a national library, consequently. For example, in Canada, national archives were founded in 1872 as the Dominion Archives. The first national library was established in 1953 as the National Library. These two institutions remained separate for many years, until 2004, when they were joined as Library and

[5] These rules would eventually lead to the Anglo-American Cataloguing Rules, Second Edition (AACR2) in 1978, and the Resource Description and Access (RDA) in 2010.

Archives Canada, reflecting the similar objectives of both institutes, and the growing bridge between them in the digital age.

Public libraries became firmly established in both Europe and North America in the twentieth century. The image of libraries as circulators of materials, and archives as preservers and collectors of unique and historical sources was well instituted by this time. However, the joint history of library and archival practices allow for influence between the two institutions (Hedstrom & King, 2003; Summit, 2008). Further, the neat distinction between libraries and archives was actually more blurred than commonly portrayed throughout the twentieth century.

Intersections of Archives and Libraries

Although libraries and archives are often thought of as a binary, a more accurate portrayal of knowledge institutes is a spectrum, with modern public libraries on one end, and modern archives on the other (Hedstrom & King, 2003). In between these two ends lie many institutions with elements of the two extremes. Collections along the spectrum may comprise diverse materials, including books, manuscripts, rare books, maps, archival materials, and serials (Hedstrom & King, 2003), and all share similar goals surrounding collecting information, and imparting it to others.

Upon this spectrum, special collections libraries fall closer to archives than any other kind of library,[6] including academic libraries, largely due to the unique materials contained within special collections. These materials include not only rare books and manuscripts, but papers or other items that could also be theoretically held in archival institutions. Like in archives, these materials generally have restricted circulation.

---

[6] The Online Dictionary for Library and Information Science actually cross-references their definition of special collections libraries with archives (Reitz, 2014d).

The development of provenance theory in special collections is therefore tied to its development in archival institutions. Beyond the nature of the two institutions, it is clear that many similarities can be drawn from historical influence. Archives began by developing their own standards, and special collections libraries in both Europe and especially the United States (where the two institutions have profound ties) were often housed with archives in universities and other knowledge institutions, especially early on. It seems natural then that provenance in special collections libraries would also be varied by standards. In fact, most libraries use some mix of national standards and local interpretation of the concept (Lundy, 2008).

This makes sense, as early archival and library schools often trained caretakers for both institutions together. As well, both archives and special collections became important for national identity in both Europe and North America towards the nineteenth and twentieth centuries, allowing the two to influence each other on institutional levels.

If provenance was not a mandate of public libraries, whose primary role is the circulation of popular materials, and not the collection of unique ones, and provenance never played the same foundational role in libraries that it did for archives, it seems natural then that provenance would be introduced into international cataloguing standards for libraries at a later point in time. Until interest was drawn towards provenance, it was not readily considered in library standards. Special collections may have utilized their own methods based on archival institutions before then.

Multiple guides currently exist for capturing provenance in special collection libraries (see “Provenance as Metadata”), and the rise of provenance in libraries coincides with the release of English translations of archival manuals, which detailed provenance

theory (particularly the English translation of the Dutch Manual in 1940, although the concept of provenance had already been applied in English archives before then). Similarly, interest in provenance in book history from the 1970s through to the present grew due to increased focus on the historical context of the works within special collections – this focus on context echoes in the development of provenance in archival theory during the same time period.

Looking forward, links between archives and special collections become only more apparent. As more manuscripts and rare books become digitized, institutions face the problem of increased amounts of data, that may be difficult to sort, display, and understand. Archival institutions historically dealt with this issue using provenance theory, and provenance could play that role again in special collections. Further, in the digital realm, archives are focusing more on the contextual nature of records and provenance. This focus will affect library holdings too, especially as more data becomes available to researchers online.

Unlike archives, provenance never became a foundational principle in libraries, as can be determined by the numerous cataloguing and classification standards in which provenance plays only a small part, if it is even considered at all (see “Provenance as Metadata”). Due to the diversity of libraries and their evolution, as well as the complex nature of provenance information itself, the growth of provenance in libraries is complex. To that end, it is useful to look at the development of provenance theory within libraries, as it is somewhat separate from the overall history of libraries.

Provenance Theory in Libraries

Interest in book provenance has always had highs and lows (CERL, 2014). Current definitions in library authorities remain rather broad: the *ALA Glossary of Library and Information Science* describes provenance as "information concerning transmission or ownership"; *Harrod's Librarian's Glossary and Reference Book* (10th edition, 2005) elaborates on provenance as also including "special binding, bookplate, or inscription that might indicate previous owners, collections, or libraries through which a book has passed"; and the *Online Dictionary for Library and Information Science* also describes provenance as being a record of ownership of copy-specific materials, and provides examples of provenance evidence (Reitz, 2014b). These definitions likely remain purposefully vague so various cataloguing standards can be accommodated through the definition.

Although recording provenance information in books seems to date back to the origins of book-collecting (Buchanan, 2011), particularly in the form of marginalia, libraries seem to have captured provenance beginning in Ancient Egypt in the third century BCE (see "Provenance as Metadata"). More traditional marks of provenance began to appear in Europe starting in 800 CE, as the use of inventories, wills, household account books, diaries, and personal papers developed (Buchanan, 2011). Bookplates, another key provenance mark, began to be used in the sixteenth century, spreading from Germany to France and Italy, and growing more elaborate with time (Buchanan, 2011).

Although private collectors often kept track of their own collections' provenance for matters of value and authenticity, libraries generally did not actively collect provenance information between the sixteenth and eighteenth century. In the eighteenth

and early nineteenth century, however, interest in provenance grew because of the value of books associated with "great men" (CERL, 2014), and possibly due to influence from archival theory, the term provenance fell into more extended use in many disciplines, including special collection libraries. Some private collections even used provenance to help organize their collections by shelf, similar to how archival institutions were organizing their materials, although it was not necessarily recorded as data at this time.

When card catalogues were introduced into libraries at the start of the twentieth century, provenance moved beyond item location on shelves to being part of catalogue records (Buchanan, 2011). According to methods based on local practice and suggested cataloguing standards, provenance information was captured in some catalogues, and often kept in bibliographic notes (Buchanan, 2011). For the most part, however, libraries gave little or no attention to provenance information for the first three-quarters of the twentieth century (Pearson, 1997), and indeed, many libraries today continue to focus on text over copy-specific information in catalogues (Pearson, 1997).

By the mid-twentieth century, interest was rising gradually in rare book studies. The Association of College and Research Libraries (ACRL), a division of the American Library Association (ALA), created the Rare Books and Manuscripts Section (RBMS) in 1948, to support special collections communities and issues relevant to their collections (Reitz, 2014c). Following this development, provenance was defined in the authoritative *ABC for Book Collectors*' first edition in 1952. This initial definition was three paragraphs long, and included some methods of assessing provenance (Pearson, 2005).

Provenance truly landed on the radar of both scholars and librarians in the 1970s (Bruni, 2011) after the lecture given by Frederick B. Adams in 1969. Adams was the

Pierpont Morgan Library director in New York between 1948 and 1969, and drew on much of his talk from the world of museums and art, where provenance theory is thought to have originated. His lecture, *The Uses of Provenance*, was one of the first discussions of provenance theory for rare books and special collection libraries (McQuillen, 2013).

Adams discussed provenance evidence, including armorials, names, monograms, re-bindings, book inscriptions, bookplates, annotations, catalogue records, and contemporary publications (Adams, 1969). Also explored were potential uses of provenance (Adams, 1969). It is worth noting that the rise of scholarly interest in provenance theory for rare books coincided with a renewed interest in archival scholarship, surrounding the context of archival materials. Although the official definition of provenance lacked the particular meaning it had in archival institutions, provenance was evolving in both archives and special collections for similar reasons.

In the 1980s and 1990s, disciplinary studies related to the history of the book continued to expand, and by extension, provenance also continued to develop (Buchanan, 2011; CERL, 2014). Advancing beyond the simple custodianship of important people, provenance began to be important for assessing the history of reading and literacy, and its cultural impacts. Any ownership, even from unknown figures, was significant (CERL, 2014; Pearson, 2005). Provenance evidence began to include evidence of how people interacted with books and circulated them. It could also show how books affected literary authors (Adams, 1969), and ideological beliefs of certain time periods.

Scholars also began to realize that ownership itself changed over time – collections changed in sizes, language, and subject (CERL, 2014).[7] Books originated from different places at different times, showing larger trends of literacy and book economy (CERL, 2014). Provenance also provided library collections with increased security and authenticity (CERL, 2014).

In the 1990s, large developments occurred for provenance and rare books. In 1992, the Consortium of European Research Libraries (CERL) was launched, which aimed to support research libraries in the accessibility and preservation of books, manuscripts, and other printed materials up until around 1830. CERL provides databases to libraries to help trace and record provenance, and has recorded increased interest in provenance from scholars and librarians alike (CERL, 2014). In 1994, David Pearson's authoritative guide on provenance, *Provenance Research in Book History* was first published. Book history scholars conceded "if we hope to understand the medieval manuscripts that we study and the manner in which we study them, we must begin by asking where they came from" (Summit, 2008, p. 2).

The start of the twenty-first century marked a height of provenance theory (Pearson, 2005). In 2004, the release of the eighth edition of the *ABC for Book Collectors* expanded their discussion of provenance to six paragraphs, and made the notable change to highlight that provenance is "always of interest" and importance (rather than only "of interest") (Carter & Barker, 2004; Pearson, 2005).

[7] Subjects may inform scholars about the popularity of books at certain times. If more copies of a book existed at one time, there may be a greater likelihood of those books surviving to the present day.

Numerous new works and projects related to provenance have also been launched. Historical collections have been reconstructed, and an increased number of catalogue lists and provenance have been made available (see Rueda-Ramirez & Ruiz, 2009; Selwyn, 1997). In 2002, Heather Jackson's *Marginalia* was published, examining the markings left behind by owners and readers of books. Specific books have also been examined and researched (see Gingerich, 2005), in order to determine their individual impact on cultural, economic, political, and book history (Pearson, 2005).

The number of conferences, publications, listservs, and databases associated with provenance has also increased, especially within CERL, where provenance has steadily increased as a theme in research (McQuillen, 2013). An increased call for provenance databases has also become apparent in the twenty-first century, alongside increased online presences of libraries and digitized rare materials. These are meant to collect library provenance data from across the world.

For example, CERL has created the Material Evidence in Incunabula (MEI) database to "record and search the material evidence of copy-specific, post-production evidence and provenance information" (MEI, n.d., para. 1). The University of Toronto hosts a British Armorial Bindings database, in conjunction with The Bibliographical Society of London (University of Toronto, n.d.). Such databases have been hailed as "integral to the future of rare book studies" (McQuillen, 2013, p. 125), in combination with continued coverage of provenance details in rare book catalogues, and provenance research by institutions and researchers.

However, Pearson (2005) notes that "many tools we have at present [for the display and recording of provenance] are dated" (p.9), perhaps because software or code

have not been maintained, or the tool design itself has become old-fashioned. In response, both libraries and archives are seeking new tools to enhance provenance information.

In spite of the growth of interest and research in provenance theory, some guides suggest provenance fervor is ill-warranted. The second edition of *Descriptive Cataloging of Rare Book*, published in 1991, advises that provenance should only be captured if important or of interest (Sweeney, 2008). The 2004 edition of *ABC Book Collectors* reinforces the idea that too much time is spent assessing provenance, and suggests, "the generally laudable attention paid to provenance […] is occasionally pushed to a length which, if not in itself slightly ridiculous, has in recent years begun to be indiscriminate" (p. 180). The collection of provenance data in libraries is still not universal, compounded by the fact that it has been estimated that as many as half the books contained in special collections contain no provenance information (Pearson, 2005)

Provenance has become a topic of renewed interest in libraries however, due to digital mediums. Libraries, archives, and other knowledge institutions such as museums are gathering together in a way they never have before, allowing for searches, cross-linking and interactions between institutions and their data that were never possible before. As provenance data plays a part in all of these institutions, it is likely that different institutional interpretations of provenance are blending together again. At the same time, digital tools offer potentially new possibilities for capturing the complex relationships present in provenance. New explorations of provenance theories in this context also allows for a return to the blurred lines between institutions.

Now, this thesis will turn to “Provenance as Metadata” in order to assess changes that occur to provenance when it is presented through the metadata standards available to special collection libraries, as well as contemporary uses of metadata.

CHAPTER THREE: PROVENANCE AS METADATA

In addition to understanding provenance as a theory, it is also important to understand how provenance information is captured as metadata by knowledge institutions. Both these understandings provide the basis needed for assessing the current state of provenance metadata. In the previous chapter, provenance is explored as complex in nature, and composed of multiple axes (owners, place, and time), that also have multifaceted relationships with each other. Its display in metadata format has implications for how the complexity of provenance is or is not understood and used by researchers.

This chapter turns towards examining metadata, its contemporary uses, and the interpretation of provenance in many metadata standards and schemas.

An Overview of Metadata

Often referred to as "data about data", metadata is structured data that can be used as a surrogate for a resource (Kennedy, 2008). Metadata in some form has been used for centuries (Kennedy, 2008; Liu, 2007). It serves roles in managing, retrieving, increasing interoperability, or usability of information resources (NISO, 2004). In digital mediums, metadata can determine how an object is displayed or preserved. It may also be key in creating virtual environments, and linking information together (Kennedy, 2008).

Many formats of metadata exist, with many different purposes. These have been created by different knowledge communities (e.g. museums, archives, and libraries), or by individual institutions, in order to meet their own goals and priorities, such as preservation, website design, *finding aids*, etc. (Liu, 2007). Different types of items have also necessitated different metadata (NISO, 2004). For example, museums may be concerned more about dimensions of objects, whereas public libraries may not value

dimension information in the same manner. Crossover between all knowledge institutions, however, is also common, especially where provenance metadata is concerned for special collections, archives, and museums.

In knowledge institutions, metadata is usually quite structured, guided by standards, schemas, and suggested vocabularies. Metadata standards are required elements (e.g. title, author, etc. and the format this should take, such as lastname, firstname), used help ensure correct use and interpretation of data. Metadata schemas are sets of metadata elements that have been designed for use together towards a specific purpose (NISO, 2004). Suggested vocabularies provide recommended ways of phrasing information, so that it is more consistent.

Metadata is also generally sorted into types based on the roles it was meant to play in providing greater access to information resources (Liu, 2007). As a complex theory, provenance actually reaches across many major types of metadata. It is descriptive, as assessing provenance requires some description of the item (e.g. marginalia, bookplates). It is administrative, as it deals with issues of ownership. It is structural, as one item might actually be made up of other items. It deals with rights, as provenance could establish who legally owns a work. Copy-specific information may be key in preservation metadata. Digitization can also be considered a step of provenance as part of the history of an item, which would be relevant to technical metadata.

As provenance can be considered any of these types of metadata, it stands that provenance metadata could exist in some format under any of these types. Most often, however, provenance is considered administrative in nature.

Metadata is usually either stored in a separate database, or embedded within a digital object (NISO, 2004). Institutions that use metadata standards and schemas help ensure that information can be shared with other institutions, especially ones which use the same schemas or standards. However, metadata might also be harvested for use in other applications (NISO, 2004). Knowledge institutions are creating and using increasing amounts of metadata (Elings & Waibel, 2007). Computing communities have also increased their use of metadata (NISO, 2004).

Metadata standards and schemas used can be further broken down and understood based on their functions. Some standards act as data structure and fields. These are the named units of information, or the elements of metadata (Elings & Waibel, 2007). The information that goes inside a data field are metadata known as data content or values. Production of such metadata is usually done with the help of a standard or thesaurus, which acts as a guide. Metadata can also be used to capture large amounts of information, which can then be moved using appropriate tools. This is format metadata. It is generally a specific type of encoding such as XML. Finally, data exchange metadata are those protocols needed in order to share metadata records (Elings & Waibel, 2007).

Based on this understanding, Elings & Waibel (2007) have suggested a new method of categorizing metadata standards and schemas (see *Table 2*).

Table 2

Elings & Waibel's (2007) Categorization of Metadata Standards and Schemas According to Knowledge Institution and Data Function

	Libraries	Archives	Museums
Data Fields and Structure	MARC, MODS	EAD	CDWA
Data Content and Values	AACR2/RDA, DCRM(B), ISBD(A), *Guidelines for the Cataloguing of Rare Books*	APPM/DACS, RAD, ISAD(G)	CCO
Data Format	XML, METS, Dublin Core	XML, METS	XML, CDWA-Lite
Data Exchange	OAI Z39.50 SRU/SRW	OAI	OAI, CIDOC-CRM, LIDO

Note. Adapted from Elings, M. W., & Waibel, G. (2007). Metadata for all: Descriptive Standards and Metadata Sharing across Libraries, Archives, and Museums. *First Monday, 12*(3). Retrieved from http://firstmonday.org/article/view/1628/1543.

Table 2

Provenance might be present in any of these levels, which may help account for variations in provenance metadata. For example, provenance on the level of data exchange might be more broadly described than provenance on the level of data content, which may observe copy-specific information in great detail, rather than what more broadly connects provenance of items to other collections.

In addition to metadata standards, libraries tend to use authorities, thesauri, and controlled vocabulary to create much of their actual metadata (Elings & Waibel, 2007). Examples for rare book cataloguing include the *Provenance Evidence Thesaurus*. Using controlled vocabulary helps consistency, inter-operability, and the use of appropriate terms for describing objects (Kennedy, 2008). Controlled vocabularies may also reduce costs, and allow for more rapid advances in library cataloguing.

It is worth noting that many library metadata schemas use adjuncts not included

in their core rules to describe special materials. Remember that the Rare Book and Manuscripts Section of the American Library Association (RBMS) was only created in the 1940s, and adjuncts to existing metadata standards and schemas only truly began to become prevalent after re-awakened interest in special collections. Before then, special collections might have used their own local standards (or interpretation thereof) to record provenance, they may have drawn on archival methods, or omitted provenance entirely.

Archives, unlike libraries, used *finding aids* to manage their materials. *Finding aids* are subjected to different standards and formats depending on the institution, and often use free-text metadata. These were not broadly shared across communities the way library catalogues were (Elings & Waibel, 2007), because of the individualistic nature of archives. However, after the introduction of computers in archives, the need to share holding information led to the development of systems often based off library standards (Elings & Waibel, 2007), contributing to the continued blurring between institutions.

In general, both libraries and archives are guided by the International Federation of Library Associations and Institutions (IFLA)'s *Functional Requirements for Bibliographic Records* (NISO, 2004). Museums have also produced their own metadata standards and schemas. These were initially created for art objects (Elings & Waibel, 2007), but have expanded to include other museum holdings.

Metadata can be created from scratch, a time consuming task even if done by professionals, but institutions also seek ways to automate the creation of metadata (Liu, 2007). Some tools used to create automated metadata include web crawlers, which search the web and collect metadata, or metadata harvesting, where metadata is gathered by tools as with the Open Archives Intiative (OAI)- PMH (Liu, 2007).

Crosswalks are another way of creating metadata quickly based off pre-existing information – these are essentially tables that map equivalencies and relationships between metadata formats. However, it is almost impossible for crosswalks to be able to match all metadata elements in different formats, or information is lost (Liu, 2007). In almost all cases of automated metadata, human editing is recommended or required.

It is noteworthy that unique metadata elements generally associated with a specific community (such as art and architecture) are often poorly applied by other communities who encounter it (Liu, 2007). This could help potentially explain inconsistencies in the application of provenance metadata, particularly where special collections are run by broader institutions.

Print library catalogues are increasingly considered in decline. As libraries become more integrated in digital environments, most institutions agree that changes are needed (Elings & Waibel, 2007). This not only includes making resources more accessible and user-friendly, but creating metadata of increased quality (Elings & Waibel, 2007). Similarly, in archival institutions, many guidelines are currently under revision due to the large amount of backlog institutions are facing with digitizing cataloguing information (Elings & Waibel, 2007), often resulting in less depth of information, and more breadth in order to digitize as quickly as possible.

In general, knowledge institutions are moving towards more networked environments (Elings & Waibel, 2007), meaning that different institutions and catalogue records are linked and interconnected. It has been suggested that multiple institutional holdings should be combined into one network, with links back to original institutions, rather than having users search each institution individually. The interconnections of

library holdings through digital tools is also a goal of future and Next Generation catalogues in many library systems.

In “Provenance Theory”, inherent complexities in provenance theory were explored. When examining provenance as metadata, it is obvious that it is further shaped by a complex network of metadata formats, any of which provenance could belong to and accordingly be recorded differently, as well as the interests and histories of individual knowledge communities. As a consequence, provenance metadata could be structured as in libraries, it could be free-form, as in archives, or it could be affected by adjunct guides to current standards, which may have changed over time.

Further, provenance may be interpreted differently through different schemas which encourage different goals or coverage of information, depending on the community capturing information and the purpose of that metadata. This is complicated even more by attempts to harvest or share information in digital environments, which could lead to the loss of information.

This background in metadata provides insight into some of the issues facing provenance metadata, and emphasizes how important it is to revisit provenance metadata and ensure its quality in an increasingly digital world.

Contemporary Understandings of Metadata

Although metadata has roots in library catalogues, it is taking on new interpretations and uses in today’s digital world. More contemporary understandings of metadata involve its use for the navigation of web sites (Morville & Rosenfeld, 2006), the use and design of search systems (Hedstrom & King, 2003; Morville & Rosenfeld, 2006), and the linking or sharing of information (Hedstrom & King, 2003). These changes, tied

to the advent of the semantic web and big data, have also changed the way scholars do research. Metadata is key to the digital world.

Consequently, contemporary understandings affect how metadata is captured and recorded, even within library institutions. Oftentimes, these contemporary roles overlap, adding complexity and importance to metadata. Some are explored below.

Semantic web

The semantic web is an envisioned "web of data" (W3C, 2015), which would serve as an augmentation to the existing web (Yadagiri & Ramesh, 2013). It has been described as a platform or framework that would allow data to become machine readable. In turn, this would allow data to be used in more sophisticated ways, beyond its current capacities which are largely tied to display (Yadagiri & Ramesh, 2013). This includes automation, integration, sharing, and reuse of data on multiple platforms (Yadagiri & Ramesh, 2013). In other words, the semantic web would allow machines to provide better information to users, by anticipating what data users might look for, providing more accurate information, and allowing for more complex queries. The semantic web should also allow for a variety of data (Yadagiri & Ramesh, 2013), and metadata can serve as an important framework in light of this.

Libraries are well-positioned to make use of the semantic web – already, library portals exist which allow users to use a single framework to make queries across repositories, formats, and multiple sources (Yadagiri & Ramesh, 2013). Increased use of ontologies to annotate or tag information will further help libraries to share and re-use resources (Yadagiri & Ramesh, 2013). Metadata is important in this regard, and standards such as Resource Descriptive Frameworks (RDF), Extensible Markup Language (XML),

Resource Description and Access (RDA), and the use of hypertexts have all been used in libraries to this end (Yadagiri & Ramesh, 2013). When thinking about use of metadata in contemporary terms, the semantic web should be considered as libraries move forward in developing semantic web cataloguing.

Any study examining metadata should also be aware of challenges libraries and other institutions face with the semantic web. The semantic web is challenged by the plethora of metadata standards currently in existence, which may not always translate well with each other (see "Interoperability" below). In fact, unifying layers of data in a logical fashion is a continuing area of research for the semantic web (Yadagiri & Ramesh, 2013). Other challenges include the time-consuming and costly process needed to revise information to expand it to the semantic web, and that certain concepts in the semantic web may overlap, be ambiguous, or contradictory (Yadigiri & Ramesh, 2013). Most of these issues expand beyond the scope of this thesis; however, it is important for libraries to consider creating clear semantic web related policies.

When considering provenance metadata in light of the semantic web, it must be acknowledged that provenance metadata is also subject to variety in metadata. However, in line with movement towards the semantic web, information should be machine-reader friendly, and, as in library portals, ideally there should be a way to query provenance metadata across multiple platforms from a single interface. Metadata is significant for accomplishing this, and a forward-thinking tool should comply to semantic web goals, and allow for complex queries and the re-use of data.

Linked Data

Linked data is often linked to the semantic web. While the semantic web refers to a "web of data", linked data itself refers to the set of best practices of using web technologies, as recommended by Tim Berners-Lee, in order to connect information or structured data that are related to each other, and making that interconnected data available on the web (Hallo, Luján-Mora, Maté, & Trujillo, 2016). With linked data, data and objects are tied together, instead of only textual information (Binding & Tudhope, 2016; Zaino, 2013). Linked data is typically assessed using a five-star model which includes use of open standards, machine-readable formats, and linking to external sources (Hallo et al, 2016), although a detailed overview of this model extends beyond the scope of this thesis.

Coyle (2013) has posited that, "the move towards Linked Data will be the most significant change in library data in these two centuries" (in Zaino, 2013, para.1). One possibility available to libraries through linked data is the use of one common authority set of controlled vocabulary that all libraries could link to, instead of having data repeated (in various formats) across every record in each institution (Yadagiri & Ramesh, 2013; Binding & Tudhope, 2013). In fact, authority records could become associated with numerous different datasets which may be otherwise unknown to the authors of the authority records themselves (Binding & Tudhope, 2016). The use of a common ontology would be important in this model.

These practices would transform the way libraries catalogue (Zaino, 2013), allowing for more powerful semantic searches and tools, such as visualizations, to be built on library metadata (Hallo et al, 2016). Interoperability between collections could be

augmented with linked data through these procedures, or by mapping different vocabularies (Binding & Tudhope, 2016) (see "Interoperability" below), allowing broader use of data, and exchange of information.

Linked data also has the potential of creating connections with external data sources, such as through the Internet of Things (see "Big data" below), allowing for complex queries to be answered more completely (Hallo et al, 2016).

Many libraries have already engaged with linked data in some capacity. Frequently cited examples include the Europeana Data Model, the Library of Congress, the British Library, as well as WorldCat, a global online catalogue managed by the Online Computer Library Center (OCLC) (Hallo et al, 2016; Zaino, 2013). Software often associated with libraries such as Evergreen, Hydra, and Omeka, have also begun to automatically create linked data (Hallo et al, 2016). These initiatives have already reported benefits for libraries and their users, such as increased data visibility, and more versatile use and reuse of data (Hallo et al, 2016).

Challenges presented by linked data for libraries include difficulties moving towards linked data standards, which are often seen as unable to accommodate the richness of current metadata (Hallo et al, 2016; Zaino, 2013), and the resource intensive nature of creating linked data (Binding & Tudhope, 2016). However, one of the major challenges to both linked data and the semantic web is that different vocabularies are often used for the same metadata, creating complications in the mapping process (Hallo et al, 2016) (see "Interoperability" below). Notably, metadata in the form of uncontrolled text can be difficult to map (Binding & Tudhope, 2016).

Consequently, quality metadata is important for the creation of linked metadata and building applications on that metadata (Binding & Tudhope, 2016), including accuracy in mapping, appropriate use of external resources, and good documentation of the methods used to create linked metadata[8] (Binding & Tudhope, 2016). Assessing the quality of metadata across many datasets is difficult to do. Other problems associated with linked data include lack of understanding or training among personnel on linked data, legal issues such as licensing, and lack of useful applications, and datasets for creating linked data (Hallo et al, 2016).

Looking forward, progress has been made towards including community in the creation of linked metadata to improve its quality and management, the creation for better mapping and data migration tools, and the development of policies for managing linked data and including it within library standards (Hallo et al, 2016). More tools are expected to develop building on linked data to aid with the analysis of large amounts of data, including tools for visualization (Hallo et al, 2016). Best practices are also developing to help libraries participate on the semantic web, and with linked data (Hallo et al, 2016).

The use of provenance metadata as linked data holds many possibilities. If many repositories link their metadata to a single authority record, some issues in recording provenance metadata may be solved (see "Lack of Guidance in Capturing Provenance Metadata" below), allowing information to be more easily accessed and catalogued in a consistent, structured manner. Further, metadata across many databases could become linked together, increasing data interoperability, visibility, and enabling the building of tools on that metadata. This, in turn, could allow the book history community to become

[8] This documentation has contributed to the growing importance of the provenance of data in computing science fields, as discussed in "Provenance Theory".

more involved in the creation of better provenance metadata, and allow researchers to do new kinds of collaborative work with other institutions, and academics in other fields. Linked data provides an opportunity for the transformation of the lone scholar working with specific datasets into a community addressing many needs and facets of research with sophisticated digital tools.

Interoperability

One of the more important roles of metadata in a contemporary sense is that of interoperability, due to the increased power of software and search engines. Interoperability refers to the exchange of metadata between two different systems, usually with the goal of allowing tasks to be executed across multiple data sets (Neiswender & Montgomery, 2009). As mentioned above, linked data may help increase interoperability of collections, such as through the linking of open vocabularies. Interoperability is therefore tied to both the semantic web and linked data. The exchange of metadata, essential to interoperability, ideally occurs in a way that prevents the loss of information. In order to accomplish this, standardized metadata, or metadata subjected to a crosswalk tool are recommended, in order to create consistent metadata.

Crosswalks are often associated with interoperability, although, as discussed above, metadata cannot be perfectly crosswalked. If metadata is inconsistent, then work by hand is often needed to clean and make metadata usable across many platforms (Challenges and Issues with Metadata Crosswalks, 2002). These problems play into the issues of the semantic web, challenging its development and use in library systems. The power of interoperability, however, makes it worthwhile to explore the creation of metadata and tools profiting from its potential.

As already described in “Semantic web” and “Linked data”, interoperability allows provenance metadata to be seen across multiple collections, that previously may never have been able to interact before. This allows new patterns and research to become exposed in book history.

Federated searches

Federated searches refer to the ability to search through multiple data sets at once using a deep linking capability, through a common interface (Caswell & Wynstra, 2010). This process is meant to streamline the search process for the user, so that retrieval and discovery of relevant information is faster and easier to do. Federated searches are also described as cross-database searches (Caswell & Wynestra, 2010). They can be traced back to the early 2000s, and often the Google search is suggested as the authority (Caswell & Wynstra, 2010).

These are increasingly used in libraries, such as library portals described above. Federated searches help users explore lesser known databases, enabling more sophisticated research, and reduce the time users need to spend learning specific interfaces (Caswell & Wynstra, 2010). The development of good discovery tools or other similar gateways are therefore important to federated searches, although this can be challenging to do. The creation of such a platform for provenance metadata could be greatly beneficial to scholars who must otherwise perform multiple searches across institutions.

Quality metadata is important for federated searches, in order to help refine results. Metadata used in federated searches can include harvested metadata, standardized metadata (through the use of controlled vocabularies or thesauri), or metadata from

multiple schemas (Woodley, 2008). As mentioned above, connecting provenance collections through linked data and associated tools can help expand federated searches, making rich discovery easier for users.

Big data

Technology has allowed for the creation of massive amounts of structured and unstructured data. These amounts are so large, that is difficult to make sense or use the data with traditional database and software methods (Beal, 2015) – this is what is referred to as big data. Big data can be a powerful tool. Although many definitions do exist for big data, the "4V" characteristics are emphasized by the International Data Cooperation (IDC) (Anagnostopoulos, Zeadally, & Exposito, 2016; Chen et al, 2014). These characteristics include volume, variety, velocity (the speed at which data is created), and value (the data has high value, but low density) (Chen et al, 2014). The "4V" characteristics lend importance to big data, making it comparable to material assets and human capital (Chen et al, 2014).

The understanding of big data has evolved over the past twenty years (Chen et al, 2014), as the size of data, much of it unstructured, began to exponentially grow due to the development of the web, the cloud, and the Internet of Things (IoT)[9] (Anagnostopoulos, et al, 2016). To give some idea on the rapid expansion of data, the IDC estimates that the size of data doubles every year (Anagnostopoulis, et al, 2016), and in 2011, the estimated amount of data was 1.8Z, and it was further estimated that data had increased close to nine times over within a five-year span (Chen et al, 2014). For the humanities, big data includes information in the format of images, manuscripts, maps, videos, digitized music,

[9] The IoT looks at how real-world objects can be connected in ways that exchange information and allow those objects to perform tasks together (Chen et al, 2014).

texts, transcripts, etc. (Williford & Henry, 2012b). Focus has shifted in many disciplines from preserving data towards finding productive means of using big data – rather than technology driving advances, information itself does (Chen et al, 2014), especially since big data provides opportunity to analyze information that would have taken lifetimes to study before the digital age (Williford & Henry, 2012b).

New kinds of research have opened up to humanities scholars because of big data. One of the most cited advantages of big data is the ability to analyze and identify trends or patterns within data, and to discover new correlations between data, rather than focus on causal relations (Anagnostapoulos et al, 2016; Williford & Henry, 2012b). The identification of trends can also reveal what information lies outside of those trends. Researchers also have the ability to study certain "slices" of data both closely and from a distance, or from a new direction altogether, particularly with analytical tools such as visualizations (Williford & Henry, 2012b), by blending or comparing different analyses. The ability to identify patterns in a "larger picture" changes the kinds of questions researchers ask of data (Chen et al, 2014; Williford & Henry, 2012b), potentially opening new directions of study. For example, Williford & Henry (2012b) present a case study from the *Digging into Image Data* project, where digitized images of historic maps of the Great Lakes between 1650 and 1800 were compared. This comparison showed inconsistencies in lake borders, leading the researchers to question whether or not the maps were actually changing due to water-level fluctuations caused by ice cover.

In addition to new avenues of research, big data promotes the use of analytical tools such as large-scale tables, and data-oriented software (Chen et al, 2014). In turn, this encourages researchers to develop new skills, or to engage with interdisciplinary

research between humanists, data scientists, and computer scientists. The traditional work of the lone scholar is therefore transforming and expanding in light of big data, towards blended disciplines and project-oriented work. The development of analytical tools, such as visualizations, are also important, in order to display analytical results so users can productively use them (Chen et al, 2014). These advances emphasize the importance of transparent and open source software (Williford & Henry, 2012b).

Big data has a lot of potential to add to humanities research, although some notable challenges exist. The most significant is the unstructured, heterogeneous nature of humanities data, which may require a lot of manual "cleaning" to be useful and reusable, particularly for traditional systems which use relational databases (Anagnostopoulos et al, 2016; Chen et al, 2014; Williford & Henry, 2012b). This cleaning many include not only standardizing data, but adding data, updating out-of-date data, converting formats, etc. In fact, data which is not manually adjusted can affect the results of big data analytics significantly, and many humanities scholars working with big data consider the creation of "clean" data equally important to the analyses drawn from the data (Williford & Henry, 2012b). However, other scholars suggest that errors in data should simply be accepted, even if errors become multiplied in analysis (Williford & Henry, 2012b). Because of this, data heterogeneity suggests larger issues within metadata itself. The diversity found in metadata is a potentially significant risk for technological advances. Data variation is therefore an ongoing issue in big data, that institutions must recognize.

Williford & Henry (2012b) have found that data cleanup is often considered a service rather than core research in and of itself. It is also a service that requires advanced expertise. Libraries are well positioned to collaborate with scholars in this manner,

providing tools and training in data management for scholars, and creating multi-institutional strategies for big data, data curation, and data and tool sharing.

Cooperation and communication between disciplines, data security, data life cycle management or data curation, tool maintenance, data sharing, data storage, funding, time, and widely available analytical tools are other challenges created by big data (Anagnostopoulos et al, 2016; Chen et al, 2014; Williford & Henry, 2012b). It is of note that big data also contributes to the importance of data provenance, for integration purposes in tools (Chen et al, 2014).

Big data is tied to data analytics, and although traditional data analysis relied on relational databases, due to the volume and variety of big data, other methods have developed (Chen et al, 2014). In general, analytics are aiming for scalable systems that have the flexibility to allow different programming language, and that analyze real-time data (Anagnostopoulos et al, 2016). Often cited management and analytical tools in association with big data include cloud computing, NoSQL databases, Hadoop (an open source platform on Apache), distributed file systems, R (an open source programming language and software environment for data analysis and mining), and many more platforms (see Chen et al, 2014), which continue to emerge.

Although the number of digitized manuscripts is estimated to be near 20,000 (Menna & de Vos, 2014), significantly smaller than most scientific big data, this is still a sizeable number that can be treated much the same way. As a lone data set, 20,000 items with complex metadata is hard to process. Further, the number of digitized special collections is growing. Digitized manuscripts may be treated as big data, for example, if users were to engage in a comparative study with large numbers of digitized manuscripts

from numerous traditions, so new insights might be gained about manuscript culture (Davison, 2009). Metadata, or tools that use metadata, are needed to help extract this kind of meaningful value from big data. Assessment of current provenance metadata practices, and potential ways of using the vast amounts of provenance metadata productively therefore tie in to big data.

Data curation

Another growing use of metadata, particularly out of the digital humanities, is the management of information over the long-term, ensuring its reuse and preservation beyond the original community– in other words, data curation. Data curation is often tied to big data and data management (Anagnostopoulos et al, 2016; Tammaro, 2016), and is referred to in a lifecycle model, covering not only preservation, but procedures for creation, access, using metadata schemas, creating links between different sets of data, and visualizing data (Tammaro, 2016).

Data curation can therefore include tasks such as cataloguing, cross-referencing, adding value to the metadata, authenticating objects and information, archiving it, and seeking new methods of retrieval and portrayal (Tammaro, 2016; CLIR, 2014). Many of these aspects, particularly those surrounding preservation, require special metadata elements, to describe the "lineage of a digital object (where it came from and how it has changed over time), to detail its physical characteristics, and to document its behavior in order to emulate it on future technologies" (NISO, 2004, p. 2), the exact kind of information provenance provides. Data curation is essential in light of changing data formats, data degradation, etc. Both data curation and data provenance are therefore considered important for data cleaning and reuse (Anagnostopoulos et al, 2016).

Data curation is often cited as a means of cleaning data, as motivation to push towards data that can be published for the semantic web (Anagnostopoulos et al, 2016). In the context of libraries and other cultural institutions, data curation also includes supporting digitized materials, and any associated digital technologies or user interfaces that handle data (Tammaro, 2016). Increasingly, adding value to curated data also involves community feedback and involvement in metadata schemas, platforms, and controlled vocabularies, which encourages interdisciplinary sharing and re-use of collections (Tammaro, 2016). These combined roles of community involvement and data curation, particularly in the realm of metadata and digital tools, emphasize the important role of libraries as a bridge between researchers and information in the digital age. Seeking means of using provenance metadata in a way that echoes the needs of the community in new digital mediums is significant in light of this contemporary use of metadata.

Contemporary uses of metadata continue to expand, and new uses also continue to arise. These new understandings of metadata are symptomatic of the changes digital technologies have brought to information and information resources. In particular, data quality and the use of new tools stand out as key in contemporary uses of metadata. Users no longer simply browse for information; they expect interactive, and complex answers from digital applications (McGuinness, Ding, Glass, Chang, Zeng & Furtado, 2006). This is true for libraries as well, as more content becomes available in digital collections. These two elements of data quality and the use of digital tools allow institutions to meet these user needs.

Metadata in Digital Libraries

Scholarship is increasingly being conducted in digital spaces (Rimmer, Warwick, Blandford, Gow, & Buchanan, 2008). Digital libraries have also been on the rise. A digital library is a library that consists of digitized or born-digital collections, accessible through a computer (Liu, 2007; Reitz, 2014a). A fundamental component of digital libraries is metadata (Challenges and Issues with Metadata Crosswalks, 2002).

Digital libraries have significant differences from their physical counterparts in regards to metadata. In physical libraries, metadata is often structured and adheres to strict standards. In digital libraries, a variety of schemas may be used, and no dominant standards are currently apparent (Hillmann, 2008), although federated guidelines and documentations are developing. Consequently, digital libraries feature various depth and quality of metadata (Hillmann, 2008). Additional metadata is necessary for digital objects. This information can be quite complex depending on the objects being captured (e.g. digitized vs. born digital). Metadata in digital libraries must also address contemporary issues such as interoperability, and be usable in digital tools, in order to help meet user needs (Hillmann, 2008).

Metadata is essential to digital libraries, for the reasons listed above, but also to ensure access and discovery of collections (Liu, Xia, Zhao, Lou, & Zhang, 2004). Noted similarities exist between libraries and the semantic web (Burke, 2009), and metadata could be used comparably for the discovery of library objects. This has spurred strong interest in creating applications with linked metadata for digital libraries (Burke, 2009).

Special collections are often digitized, because of their unique and valuable nature, and therefore the metadata of those digital libraries come from a combination of

catalogues, *finding aids* and codicological descriptions (Davison, 2009) – the latter of which can extend to entire monographs. This practice continues today because "there is no single metadata standard that works for describing unique digital resources in all contexts" (Han, Cho, Cole, & Jackson, 2009, p. 235). Because of their rich metadata, special collections are often considered for semantic web growth within libraries (Burke, 2009). This can be complicated as metadata for special collections can be quite varied due to reasons explored above (Hill, Janee, Dolin, Frew, & Larsgaard, 1999).

Although the increased digital presence of special collections allows more users to access materials than ever before, there is an impetus for digital libraries to do more than just display digital items (Davison, 2009; Rimmer et al, 2008; Whittaker & Thomas, 2009). Users have come to expect digital libraries to make use of digital tools that promote discovery, interaction, and connectivity (Davison, 2009; Whittaker & Thomas, 2009) for vast amounts of information, in a relatively accessible manner (Rimmer et al, 2008). Digital libraries are not meant to simply replace physical ones; instead, they play a complementary role (Rimmer et al, 2008) where digital libraries provide not only methods of search and retrieval, but analytical or tools features based on presumed authentic and trustworthy physical materials.

Many of the features users expect tie into the more contemporary uses of metadata discussed above, such as the ability to access multiple collections through a single portal (Hutton, 2008). New metadata elements have been created; old metadata elements have been used more broadly (such as for linking objects and other metadata together) or have taken on new functions (such as preservation) in order to accommodate these needs (Liu, 2007). Initiatives for collaborative projects and the development of cross-searching

catalogues have also been supported (Hutton, 2008), and interoperability of data has also become increasingly important for digital libraries.

Libraries often engage new technologies encouraging interaction and the ability to make use of metadata to help organize and search materials in innovative ways. For example, the British Library's "Turning the Pages" project, which allowed users to "turn a page" when examining digital copies of materials, was a popular tool. Similar initiatives were taken on by other special collections later. These kinds of projects also bring acclaim and interest to institutions (Rimmer et al, 2008).

Libraries have also embraced social aspects of new technologies, such as wikis, blogs, and social networks (Whittaker & Thomas, 2009), although these often lack the depth of information the semantic web brings to catalogue information.

Many of these features, and new metadata standards and schemas, have been considered in context of bringing libraries into Next Generation catalogues. Next Generation catalogues seek to take advantage of the digital world, by implementing powerful and appealing discovery tools that not only help users locate and retrieve items, but also encourage libraries to take "a more active role when it comes to increasing the sphere of knowledge" (Kumar, 2008, p. 371). In fact, Next Generation catalogues are sometimes thought of as a tool unto themselves, designed to help users research, teach, and learn by allowing not only the discovery of data, but its use. As Next Generation catalogues continue to be a goal for developing libraries, features are continually introduced towards bringing these richer systems to light.

Provenance metadata therefore exists in a world as a complex entity, subject to many different methods of capture, but that ultimately should be data that can be used

productively for users in a contemporary sense and within Next Generation catalogues. An example of the kinds of features Next Generation catalogues might use include visualizations, which also have the potential to take advantage of many of the contemporary features of metadata outlined here.

Metadata Visualization

Metadata has been frequently used in digital tools, such as visualizations. Visualizations are often described as methods of representing information, such as metadata, through visual means. Visualizations are tools to help users understand and sort through vast amounts of information relatively quickly (Mazza, 2009; Araya, 2003). In fact, behavioral research suggests that an image is analyzed 60,000 times faster than textual data of the same subject (Anagnostopoulos et al, 2016). For example, a point on a map is understood more quickly than geographic co-ordinates simply written on a piece of paper. Visualizations can also help researchers uncover new information from data (Anagnostopoulos et al, 2016) and change the way users think about information (Araya, 2003). For example, a timeline may influence users into thinking about events as a continuous and connected stream, instead of as discrete dates. Visualizations are a commonly used tool for big data analytics.

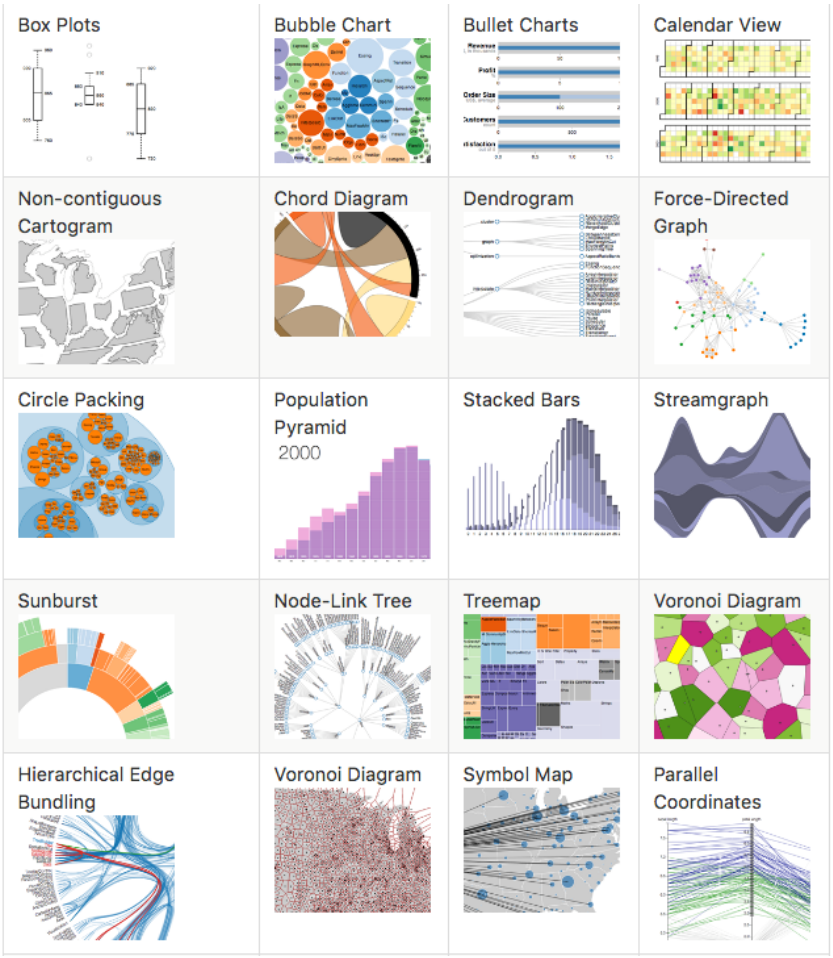

Figure 1: Examples of types of visualizations (Nylan, 2016)

Examples of visualizations include images, graphs, diagrams, maps, network webs, hierarchy maps, treemaps, heat maps, and Sankey diagrams, among others (Mazza, 2009) (see *Figure 1*). In many ways, these serve as an access point to data other than just through text (Williford & Henry, 2012a). Characteristics that have been identified in good visualizations include accuracy of data, simplicity (the visualization should not overwhelm the user), the ability to group or cluster data, and aesthetics (Mazza, 2009). Increasingly, visualizations are also seeking to indicate to users what is not shown due to incomplete data, potential errors in the data, or ways of visualizing uncertainty (Anagnostopoulos et al, 2016; Williford & Henry, 2012a). Visualizations should

additionally meet user needs by being presented in ways that are easily accessible through understood metaphors (Anagnostopoulos et al, 2016). A good visualization also helps data cleaners and analysts by helping make clear the potential uses of data (Anagnostopoulos et al, 2015). As a tool of big data, visualizations allow researchers to identify trends, as well as what is unique about a data set. As big data continues to evolve, visualizations and affiliated analytical tools will have to keep pace.

More and more visualizations in the digital realm are tending towards interactive tools, that allow users to make use of different facets or even types of visualizations at the same time (Williford & Henry, 2012a). For example, visualizations that allow users to explore collections through space and time at the same time allow different dimensions of the collection to be available to the user at once, exposing different analyses from those facets or incomplete data from one facet in the other (Williford & Henry, 2012a). Using different visualizations alongside each other also helps bridge the gap between the changeable, exploratory form of humanities, and the more rigid formatting of computational analytics (Williford & Henry, 2012a).

Visualizations therefore meet all of the contemporary needs of metadata described above. Multiple items can be combined and linked together in a single visualization, as required by linked data and the Semantic Web. This ability for collections to communicate within the same visualization requires some level of interoperability of metadata, particularly as generally metadata used in any visualization needs to be cleaned and standardized (Van Verchum & Pugin, 2014), and supports search functions across multiple collections. As described in “Big data”, current estimates place the growing number of digitized manuscripts near 20,000 (Menna & de Vos, 2014), and as

visualization inherently helps organize vast amounts of information, visualization may be a means of sorting this big data. Finally, by extension of the standardization of metadata required in visualization, metadata becomes more robust and more easily preserved in data curation.

Further, literature from scientific and computing fields have suggested that visualizing geospatial provenance metadata helps users perform queries on data (Göbel & Jasnoch al, 2001), navigate information, and understand its history (Ziheng et al, 2013). For example, visual feedback can help users enhance their searchers or discover effects in search results by modifying their queries (Göbel & Jasnoch, 2001).

Visualizations are not new to metadata or library systems, in fact, "visual interfaces enhanced with metadata are becoming more widespread to provide a richer representation of digital collections," (Shiri, 2008, p. 763) in part due to a push towards Next Generation catalogues. In library systems, visualizations have been shown to increase access to information, the ability to search, sort, and browse collections, and to make connections between concepts (Shiri & Molberg, 2005). In turn, provenance lends itself well to visualization, because of the inherent axes of the definition of provenance.

Drawing on the work of librarian and information science pioneer, S. R. Ranganathan, evaluating provenance of an item involves the five fundamental categories of Ranganathan's faceted classification scheme, PMEST: *personality*, *matter*, *energy*, *space*, and *time* (Foskett, 2009).

Provenance engages *space* (where a rare book has been), *time* (when a rare book was in a place), and *personality* (a book is distinguished from others based on who owned the item). In some cases, *matter* (physical material of a book such as bindings), and

energy (how a book changed hands – e.g. if it was gifted or sold) also enrich an object's provenance. These facets allow the representation and visualization of provenance to be metadata-enhanced. In other words, visualization could be a richer way of representing the narrative of provenance for users, if these facets are made available. Inclusion of as many of these facets as possible should be a baseline criterion for any visualization.

The success of a visualization depends on the use of accessible metaphors – visual images that users are familiar with and understand (Shiri, 2009). Of interest to this thesis are the potential visual metaphors inherent within provenance: time (timelines), place (geographic maps), and owners (networks). Although many visualizations are available for use that deal with multiple axes, this thesis will focus on these metaphors, in part because they are inherently easy to understand and accessible to those in the field. Further, all three of these visualizations have been used individually in library systems before.

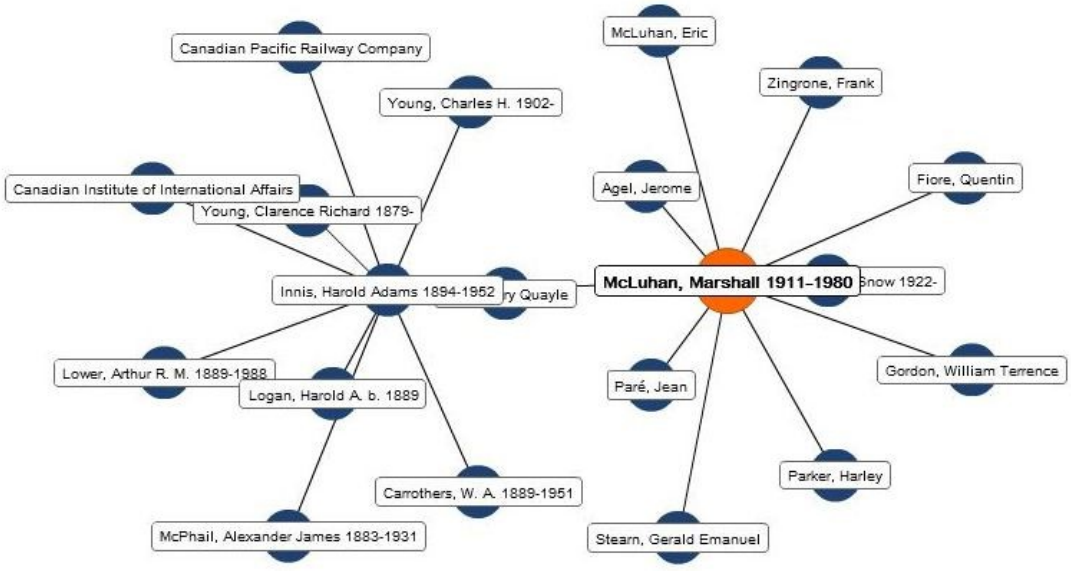

Figure 2: Network visualization from the WorldCat Identities Network of related authors to Marshall McLuhan (Richard, 2013)

Network visualizations are perhaps the most common of library metadata visualizations, and are typically used to illustrate the citation network of an item, through use of bibliographic metadata (Bergström & Atkinson, 2010). These visualizations usually consist of webs, linking different entities together through the use of lines (see *Figure 2*), although tree maps are sometimes used as well (Shiri, 2011). Both these visualizations can potentially overwhelm users by information overload (Bergström & Atkinson, 2010).

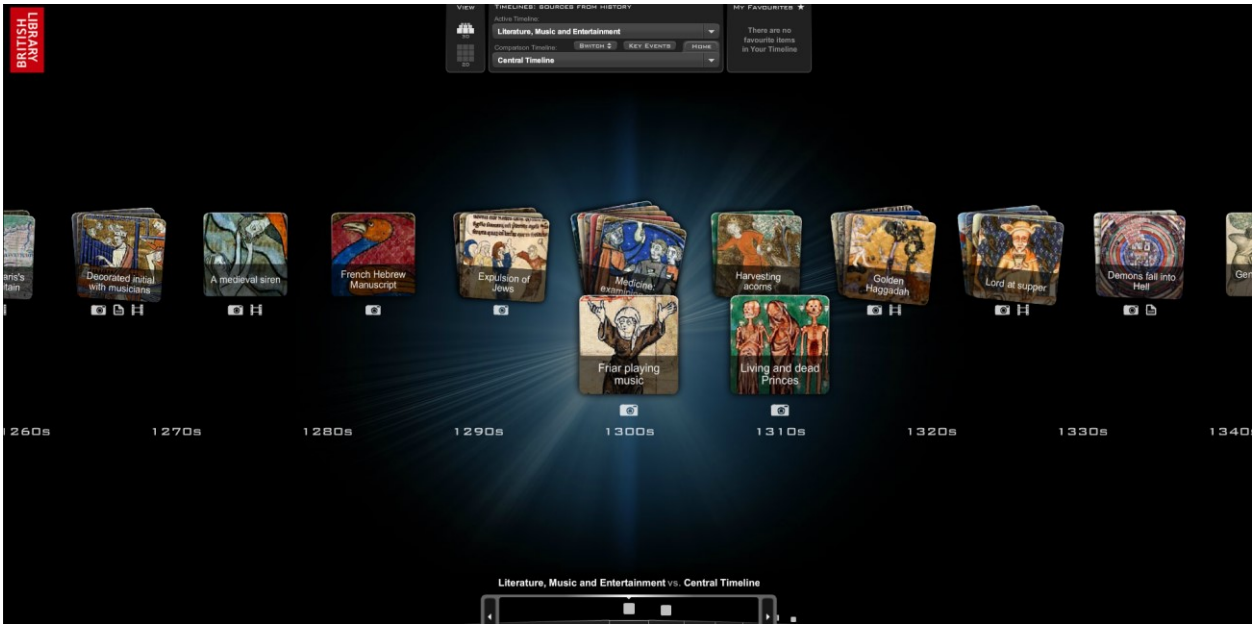

Figure 3: Timeline visualization from the British Library of medieval manuscripts (British Library, 2016)

Timelines have also been used as visualizations in libraries. Plethoras of examples are available on the web. The British Library has created numerous interactive timelines for the public (see *Figure 3*), although they are not necessarily specific to only library holdings, but instead might focus on themes like famous people or time periods (British Library, n.d. c). The Sonoma State University has also used the Massachusetts Institute of Technology (MIT)'s Semantic Interoperability of Metadata and Information unLike

Environments (SIMILE) program to successfully generate a digital, web-based timeline based on their collection's metadata (Marquez, 2012).

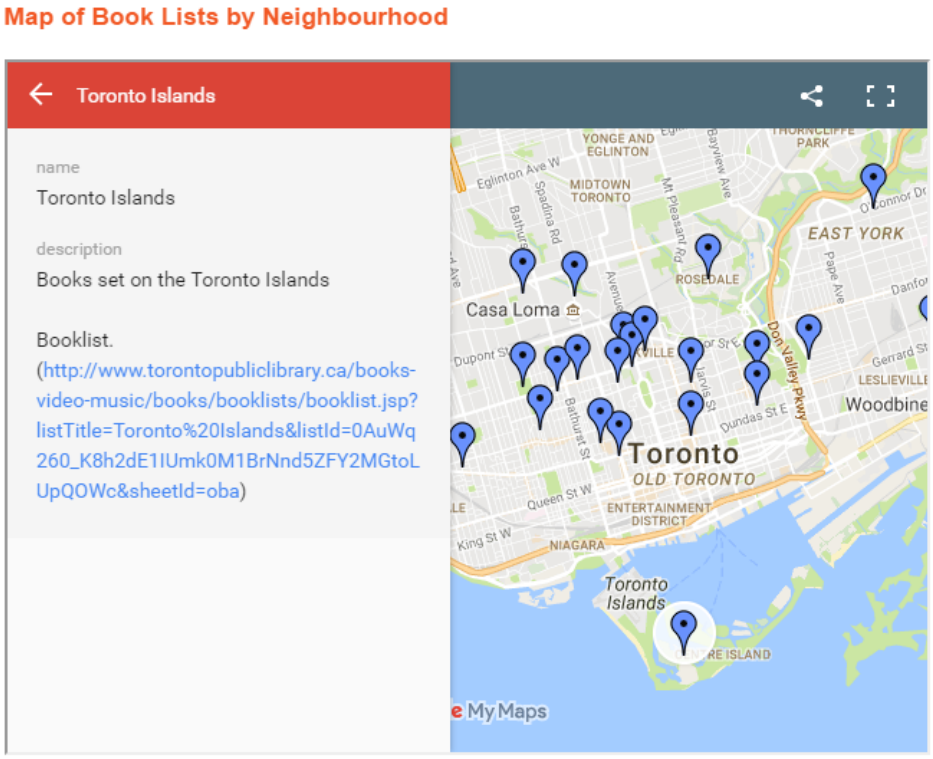

Figure 4: Map of the Toronto settings of book lists from the Toronto Public Library Collections (Toronto Public Library, 2016*)*

Recently, digital libraries have also been increasingly using digital maps for the following purposes: creating geological visualizations, digitizing collection maps and atlases, making products available to patrons, geopreservation, mapping collections themselves (see *Figure* 4), and to track information about patrons to better accommodate their needs (Deckelbaum, 1999). Digital maps can also highlight information on trade, economics, and politics of a certain time (sassen, 2008). In order to use digital maps, geographic-based metadata must be available.

Digital Library Metadata Standards

The potential of metadata to bring new methods of discovery and research to libraries and their digital counterparts is apparent. In spite of the increased role of metadata in digital libraries, standards are inconsistent. Although many guidelines for digitizing materials are in place, no standard formats have been selected for digital libraries (Liu, 2007). Liu et al (2004) note that "practitioners are always suffering from no guidelines or pilot projects [regarding metadata in digital libraries that] can be referenced when developing digital library applications" (abstract).

As a result, crosswalks between systems often yield poor metadata, and digital tools and searches become more difficult to run. Currently, metadata is often repurposed or enhanced in order to be useful in the online environment. Liu (2007) suggests that digital libraries require metadata standards, equivalent to AACR2 for print catalogues, and standard frameworks for metadata exchange, equivalent to MARC.

In light of the lack of standards and schemas applied in digital libraries, it is useful to turn towards examining what metadata standards and schemas are available to collections for capturing provenance, so that a better understanding of what might be used in online environments can be gained, and how varied those metadata elements are.

History of Cataloguing Provenance

Although the history of provenance as a theory within libraries has been discussed in "Provenance Theory", with a better understanding of metadata and how provenance interplays with metadata, it is possible to look at a history of provenance in libraries as metadata, and not just as a theory.

Library catalogues have a longstanding history, dating back to the Alexandrian Library in the third century BCE (Taylor, 2006). This library was also the first to capture provenance metadata, through a note reading "from the ships" in the catalogues, for those manuscripts which were confiscated from vessels going through port and added to the collection (Kennedy, 2008). Aristotle's private library was the first whose history could be traced (largely since the collection travelled as a whole) (Buchanan, 2011), and the Romans may have also actively captured provenance data (Posner, 1972).

Over the following centuries, recording provenance was not consistent in libraries, although it has been documented in books as long as books have been collected (Buchanan, 2011). Provenance first seemed to be used as a means of establishing value for rare books in the eighteenth and nineteenth century, but it also was used to help organize some special collections in early cases. As cataloguing systems and metadata standards developed, issues of provenance were generally addressed by recommending the simple recording of information in card catalogues (Lundy, 2008).

Provenance experienced a re-emergence as a theory of interest in book history in the mid-twentieth century, however, and the concept continued to grow in interest beginning in the 1970s. In spite of this, cataloguing instructions regarding provenance remained vague. Most catalogues only made note of previous owners with no regard to additional context. Pearson (1997) observed that by the end of the twentieth century, some fifty years since provenance theory began to develop in libraries, that

> most of the catalogues of historic libraries created anytime during the first three-quarters of this century or earlier give scant or no regard to provenance matters,

> and there is plenty of cataloguing activity going on today which still regards the text as the only thing worth describing in a record. (p. 505)

Near 67% of a sample of 248 American rare book collections did not provide provenance information, according to a study conducted in 1996 (Overmier & Doak). More recent assessments have not been made, and certainly not in digital mediums.

One of the most significant events in the history of cataloguing provenance metadata occurred in 1979, when the Independent Research Libraries Association (IRLA) released a report examining the cataloguing of rare books, and associated metadata fields. The report found that current standards were lacking and inconsistent. This caused concern, as computerized catalogues were opening up new possibilities for libraries, and local bibliographic information became available to other institutions (Lundy, 2008).

Consequently, new fields were introduced for capturing provenance metadata in the MARC format. The standard Bibliographic Description of Rare Books (BDRB) was created in 1981 to supplement Anglo-American Cataloguing Rules, Second Edition (AACR2), as well as new thesauri such as the *Providence Evidence* thesaurus (1988), and other similar tools such as the *Printing & Publishing Evidence Thesaurus* (1986) and the *Binding Evidence Thesaurus* (1988). These developments reflected an active effort from the part special collections communities to standardize metadata and terminology. The 1979 IRLA study is considered "the real beginning of the focus on developing standards for rare book cataloguing in the online environment" (Lundy, 2008, p. 166).

Throughout the 1990s, new emphasis was placed on recording copy-specific information in catalogues, and including detailed records of provenance for both preservation and research purposes (Lundy, 2008).

Researchers claim that the capturing of provenance metadata has greatly improved since the turn of the twenty-first century (Pearson, 2005), although no statistical evidence from comparable studies to Overmier & Doak's (1996) currently exists to support these assertions. Catalogue instructions for materials now include more complete information featuring expanded discussions of dates, places, and relationships to other materials. One of the bigger roadblocks remains backlog and conversion of old records into new formats (Pearson, 2005), both of which are time consuming and may require revisiting the object(s) in question and navigating multiple metadata schemas and standards. The latter are also frequently being updated – for example, the Descriptive Cataloguing of Rare Materials (DCRM) provided more copy-specific guidelines for capturing provenance in its most recent edition. Some tools, however, are still in need of updating, such as bookbinding databases (Pearson, 2005), which may be lacking information or no longer be up to current standards.

An examination of the current recommended standards and schemas follows here.

Provenance in Metadata Standards and Schemas

Examining both metadata standards and metadata schemas, and their interpretation of provenance into metadata is integral to understanding what provenance information might currently be available to users, and the depth and complexity of that information, as well as challenges libraries might be facing. This section explores the interpretation of provenance within major metadata standards and schemas. For a full list of standards and schemas examined, metadata elements used to capture provenance, and the scope and definition of those elements, see *Appendix A*.

Twenty-six major standards and schemas were examined, drawing from the knowledge communities of museums, libraries, and archives. Standards and schemas were also chosen from across the different function types outlined by Elings & Waibel (2007): data structure, data content, data format, and data exchange. Examining provenance as metadata across these standards and schemas revealed that many potential options are available for capturing provenance, and these are quite varied.

Museum standards, surprisingly, do not always include an element for provenance metadata, but where elements are present, they tend to be quite comprehensive in scope. Library standards, on the other hand, are most likely to use broad categories such as "Notes". Archival standards are most likely to provide multiple fields for capturing provenance, which aim to be very comprehensive in scope. These findings make sense in light of the history of provenance theory in libraries and archives (see "Provenance Theory").

When examining provenance metadata elements according to function type, more variations become obvious. Metadata standards traditionally associated with data structure all have provenance fields, which tend to be fairly comprehensive in scope. Broad categories such as "notes" to capture metadata, however, are most overwhelmingly found within data content or descriptive standards, which are the most likely to be displayed to end-users. Archival standards that fall within this category are exceptions, however. Metadata standards which fall under the function of data format are most likely to capture provenance using a scope that is digital in nature, and is therefore more narrow in definition. This trend makes sense, considering the role of data format. Finally, those metadata standards more traditionally associated with the function of data exchange seem

to have no cohesive trend, and can vary from very narrow scopes to very broad scopes, likely contributing to the challenges of using crosswalks on provenance metadata.

In total, five standards offered no means of capturing provenance information. In two standards, provenance is only accounted for digital items, and not physical ones. Eight metadata standards break up provenance metadata across multiple fields. All standards with provenance elements request the capture of ownership information; however, not all request contextual information such as dates. Further, some standards specify only capturing information relevant to the most immediate owner, instead of a total history of ownership. The depth of information within a provenance element therefore also varies by standard.

Finally, the terms used to describe provenance can vary, including not only Provenance, but also Custodial History, Former Owners, or Source of Acquisitions, among others, for a total of forty-two different elements (see *Table A1*).

From examining current metadata standards, it is apparent that available provenance information is not consistent in presentation, type of information, or depth of information. This plethora of metadata standards seems to exist for many multifaceted reasons, including the complexity of provenance as a theory itself with many axes, and the complex historical development of provenance in archives as a definite, concrete theory, and in libraries as a more general add-on. Different knowledge communities thus seem to focus on different aspects of provenance when capturing it, and this is further complicated by the different functions of metadata, and more contemporary roles of metadata. Examining current standards, however, also suggests that capturing provenance may be challenging due to the guidelines themselves.

Lack of Guidance in Capturing Provenance Metadata

Although provenance theory has experienced a re-emergence in book history, and the evolution of provenance theory has led to increased capture as metadata, the literature agrees that on top of the plethora of metadata standards and schemas with different interpretations of provenance, many do not provide good direction for recording copy-specific information (Pearson, 1997). In fact, for many years, "one of the problems in the English-speaking world […] has been the lack of guidance within the cataloguing codes" (Pearson, 1997, p. 505). This has been referred to as a "major [shortcoming]" (Pearson, 1994, p. 318).

Further, provenance information consists of many facets. Unlike other metadata fields such as title or author, which are fairly similar across standards, provenance can be hard to confine to a single metadata field, as it includes not only owners, but contextual information such dates, places, methods of transfer, or provenance evidence on the actual object. As a result, many libraries may provide only the name of a known owner, meaning that other facets are unaccounted for, or would require extensive research to fill. Other times, provenance is simply spread out over many metadata fields (Buchanan, 2011; Greenberg, 2005; Lundy, 2008).

Another challenge cataloguers may face when creating provenance metadata occurs when attempting to use standard vocabulary. Although many thesauri exist to help cataloguers, not all owners or terms may be covered in such thesauri, or issues of translation may also obscure the use of standard vocabulary. Other times, provenance evidence such as handwriting on an item may be illegible. Lack of common thesauri and

vocabulary also cause problems in recording provenance metadata, as many collections opt not to use existing resources.

As recently as 2007, the RBMS MARC for Special Collections Discussion Group at the American Library Association Midwinter Meeting made note that although many standards exist for recording provenance, even when a standard is chosen, most institutions just relegate the information to added entries or local fields (Lundy, 2008), leading to unstructured notes that are not useful for scholars. All of this can make identifying the provenance of an item difficult for scholars working at a distance. The discussion meeting concluded that many libraries use a mix of national standards and local interpretations of those standards, in order to create their own practice of recording provenance (Lundy, 2008).

Ideally, cataloguers would be able to consult not only the items themselves for provenance information, but could cross reference other documents such as inventories. Most often, however, cataloguers simply do not have the time or financial resources to spend creating a full and formal history of every item in the collection (Bruni, 2011). Working in such constraints is a major challenge for special collections, which can also lead to provenance metadata being omitted or included with less depth.

Although it has been posited that progress has been made in recording provenance metadata (Pearson, 2005), very few studies have examined or quantified this progress, or the current state of provenance in library catalogues – particularly in digital libraries, which already struggle with metadata standards. “Methodology” will look towards the methodology and outline of a two-part study of provenance metadata: first, to assess current use of provenance metadata in digital collections, and second to examine the

potential of visualization tools to address some of the limitations regarding provenance metadata, while also providing libraries with a Next Generation tool.

CHAPTER FOUR: METHODOLOGY

The literature has shown that provenance is a complex theory, with an equally complex institutional history. Multiple standards and definitions exist for capturing provenance as metadata. Researchers in the 1980s and 1990s suggested that there was little guidance in cataloguing this process (see Buchanan, 2007; Lundy, 2008; Pearson, 1997), and consequently, many libraries created their own local interpretations of provenance. Even within existing standards, guidance can be broad. More recently, the literature suggests library catalogues have improved practices in capturing provenance (see Pearson, 2005) with the help of more developed guidelines.

The current state of provenance in library catalogues remains largely unexamined, although provenance theory is experiencing a surge of research and interest (McQuillen, 2012; Pearson, 2005). The unknown state of provenance metadata is particularly true for digital special collections, an important resource for scholars. This is surprising, considering special collections also tend to lead digitization projects.

This chapter outlines the methodology for a proposed two-part study to address the current use of provenance metadata, and how best to portray that data to users with the digital tool of visualization. Through this investigation, limitations in provenance metadata suggested by the literature (see Buchanan, 2007; Lundy, 2008; Pearson, 1997) can be both verified and potentially improved upon.

The overall methodology of the full study is addressed first, followed by individual outlines for each part of the study. The first half of the study is focused on current use of provenance metadata in digital libraries. The second half pilot tests

visualization of provenance metadata. Each outline addresses the thesis research questions proposed in the *Introduction*, as well as research design, and limitations.

Overall Study Methodology

The epistemological stance of this study is modified dualist and objectivist. Under this epistemology, critical realist, relativist, and pragmatic ontologies are adopted. These ontologies recognize that provenance metadata must be tested, and analyzed in order to determine its best means of display. They also recognize that context is essential in analysis. As in similar metadata studies, pragmatism allows for examining a specific and complex metadata element that has multiple interpretations such as provenance, but also addresses the most important meanings of that metadata element (Bayih, 2010).

This thesis uses analytical induction to build an understanding of provenance by examining facts, evidence, and core characteristics of the theory (Gorman & Clayton, 2005; Wildemuth, 2009). The overall methodology is qualitative, aiming to provide a multi-layered exploration of provenance theory and use within special collections.

The overall study method may be classified as action research. In action research, current practice is not only studied to see how well it meets user needs (Gorman & Clayton, 2005), but an attempt is made to improve the situation. In this case, the current use of provenance metadata in digital special collections is assessed, and visualization tested as a means of addressing limitations. Results are also analyzed to create a set of recommendations (Gorman & Clayton, 2005; Kumar, 2011; Pickard, 2013).

Action research is characterized by its use of multiple research methods and its use of data from multiple sources (Pickard, 2013). This study can be divided into two parts, which use different methods. The first part examines current use of provenance

metadata in digital libraries in a comparative study, using bibliometric and content analysis, and light statistical analysis. This is supplemented with historical methods used in the literature review. In the second part of the study, visualization is tested using case studies or pilot tests, and analyzed using criteria based in usability and systems analysis. Due to the several methods employed, this thesis is of holistic design (Kumar, 2011).

Pitfalls of action research include researcher bias and conclusions that cannot be easily generalized, as most action research is undertaken on a local level (Pickard, 2013). This study helps counteract these limitations through an extensive literature review and use of joint methodologies that draw on multiple institutions and tools to form analyses. However, the study is focused on digital collections and does not account for print catalogues, meaning only a particular snapshot of provenance metadata is captured.

Action research is also limited by timing – certain problems may be relevant only to a certain time (Gorman & Clayton, 2005), especially in digital environments where technology advances quickly. To accommodate this issue, the study was conducted within a time limit of two weeks, to create a clear representation of information.

Study of Provenance Metadata in Digital Libraries

The methods of capturing provenance across major standards vary, not only in terms of the kind of information to capture, but the depth of which to capture information. Literature has suggested that the cataloguing of provenance metadata has many limitations in light of this, including being relegated to broad "Note" categories (Lundy, 2008; Pearson, 1994). Although it has also been suggested that cataloguing practices have improved in recent years, no studies exist to verify such information. The first half of this action research study attempts to determine the current use of provenance metadata in

digital special collections, in order to ascertain and quantify the claims made in the literature.

Research Questions and Previous Studies

At the outset of the study, a number of research questions were identified. The first two guiding questions were: what standards exist in provenance metadata, why did so many come to exist, and how do they differ; and how are special collections currently using these standards and are they meeting user needs?

The first question is largely addressed in "Provenance as Metadata", where numerous provenance metadata standards according to different disciplines and metadata categories are outlined (see *Table A1* in *Appendix A*). Some suggestions as to why numerous standards exist have been made, namely that different disciplines and communities have focused on different aspects of provenance, as informed by the history of those institutions.

Any further analysis, however, requires knowledge of whether or not the standards available to institutions are being used. This leads to the second question of the research study, assessing current use of those standards within special collections, and whether or not those standards meet the needs of scholars as outlined in the literature (Pearson, 1994).

Provenance metadata provides important information for scholars, but "Provenance as Metadata" shows metadata can also be used in a variety of ways. Already, disciplines such as archival science are pushing for new ways of portraying provenance in more complex manners. This includes the mapping of multifaceted relationships, and capturing contextual information. Meeting user needs therefore

includes capturing provenance metadata, and using that metadata in new ways, particularly in digital environments.

To be used effectively in digital environments, good quality metadata is needed. Assessing the current state of provenance metadata must therefore include discovering standards used, consistency of use, and impacts this may have on quality.

Although provenance theory is at a point of prominence within the discipline of book history, only one study, "Provenance Records in Rare Book and Special Collections" by Overmier and Doak (1996), examined the use of provenance metadata within special collections.

Overmier & Doak (1996) believed that even though numerous guides and standards encouraged libraries to maintain provenance records, most libraries were not consistent or comprehensive in doing so. Survey data was collected from 248 collections, which revealed that only 82 or 33% actively maintained provenance information.[10] Further, institutions often only recorded provenance for parts of the collection, and many marks of provenance often went unrecorded.

Factors that affected which libraries kept provenance information included geographic location, size of collections, and the length of existence of the institution. Those collections in major locations, with larger holdings, and which were more established all were more likely to record provenance (Overmier & Doak, 1996).

Roughly half of the special collections consulted in 1996 which maintained provenance records also provided that information online (Overmier & Doak, 1996),

[10] This number is quite low; however, previous studies on medical rare book libraries found that an even lower number amongst that sample collected provenance – only 15 of 70 libraries, or 21% (Overmier & Doak, 1996).

where it was believed that users could more readily access information. As suggested by literature of the time, use of standards was stratified between embedded metadata, "notes" fields, and additional entries (Lundy, 2008; Pearson, 1994). Further, within the elements themselves, the information gathered was not necessarily subjected to standard terminology. A majority of institutions used established terms, but a significant minority used natural language or terminology developed by the institution itself.

Overmier & Doak (1996) concluded that a lack of standardization in provenance metadata did indeed exist, and that such a lack of information went against most cataloguing recommendations. They theorized the lack of standardization of provenance records across institutions would impede universal access to information, and the work of scholars in the field. These findings substantiated claims in the literature at the time (Pearson, 1994).

Now, almost twenty years later, this study is out-of-date, especially considering the large changes libraries have undergone in both cataloguing and digital environments. No true focus has been placed on these new digital environments, which is the setting in which both special collections and provenance theory have seen the most growth. Considering these factors, a study on the use of provenance metadata in major digital special collections is especially timely.

Study Methods and Design

The bulk of the first half of the remainder of this thesis is a comparative study of digital special collection libraries, which seeks to determine what provenance metadata is currently used in these collections. The study is also evaluative, as it seeks to determine if provenance metadata is meeting the current needs of users. Digital libraries as defined in

this study encompass virtual collections that offer digitized items. Digital libraries have been selected for study to better understand the changing environment provenance metadata finds itself in. This study seeks to focus on information immediately available to scholars, even those working at a distance. Under these criteria, print catalogues, digitized catalogues, and collection backlogs cannot be included in this study. For these same reasons, any embedded, encoded, or markup metadata have also not been included.

As this study focuses on provenance of rare books and manuscripts, any digital library considered for study must focus on digitized items of medieval and early modern (before 1700s) European origin. Digitized items focusing on non-Western countries and other time periods therefore fall beyond this project's scope, as do born digital objects. Selected collections must also have at least two digitized rare books and manuscripts available alongside their broader catalogues.

Libraries selected for study were done using purposive or judgment sampling. Thousands of digital manuscripts exist in some format, and are available online (Echard, n.d.), and over three hundred digital libraries containing manuscripts and rare books exist (Menna & de Vos, 2015). Although this number narrows considerably once project criteria are applied, choosing an appropriate sample size is not easy. With an unknown number of collections, and such large variety, consultation of comparable metadata studies (see Park & Maszaros, 2009), and researcher assessment were necessary.

Other studies on metadata records vary in the number of records compared, from a few case studies, to sixty records across three digital collections (Park & Maszaros, 2009), or to records of hundreds of collections (Overmier & Doak, 1996). A sample size of sixty-four libraries was selected as a representative number that still falls well within

the parameters of similar studies. An attempt was made to include major digital collections suggested in the literature (e.g. if the library appeared on more recommended lists, it was considered for study).

Libraries selected were geographically based across North America, and the United Kingdom and Ireland. Only English-language libraries were selected, so that comprehensibility of metadata fields would be consistent, but also as different language libraries may use sets of standards outside the scope of this study. As the line between defining an archive and a special collection can be inexact (Duchein, 1992), only collections self-identifying as libraries were selected.

The use of known and popular digital libraries helped shape the purposive sampling used to select libraries for this study. Such collections may be more shaped towards user needs. Through consulting digital resources from projects and academic websites directing scholars towards important digital collections, including the *Monastic Manuscript Project* (Diem, 2011), the *Digitized Medieval Manuscripts app* (DMMapp) (Menna & de Vos, 2015), *Consulting Medieval Manuscripts Online* (Peckham, n.d.), *Medieval Manuscripts on the Web* (Echard, n.d.), and *Some Useful Websites in Medieval Literature and Manuscript Studies* from the University of Alberta (Reimer, 2015), a list of digital libraries was selected for study.

Following previous digital library evaluative studies (Chowdhury & Chowdhury, 2000; Kumar, Furuta, & Allen, 1998), digital libraries were classified into different categories: those digital libraries developed as part of a university (academic-based digital special collections), those digital libraries developed as a national and governmental library, those digital libraries developed by independent institutions, and

those digital libraries developed out of academic and scholarly projects. Categorization allow analysis to determine if type of institution makes a difference in collection of provenance metadata.

Where possible, allocation was made for matching libraries across geographic regions – namely academic libraries, where even numbers of libraries from the United Kingdom and the United States were selected. This helps ensure more balanced results with less bias towards any one academic system. As very few Canadian academic libraries met the study criteria, all identified were included. Similarly, as very few National Libraries exist within the parameters of the study, all were included. However, institutions falling within the "Other" category relied more heavily on reputation than geographic balance. Limiting institutions by geography would have eliminated prominent institutions from the study at hand. "Digital Projects", often collaborations amongst multiple institutions across many geographic regions, were not balanced geographically due to their sometimes inherent nature as combined ventures.

The sixty-four digital special collections selected for study are listed in *Table 3* according to type of institutions. In total, twenty-nine university and academic institutions were identified, six national libraries, thirteen other institutions (the majority of which are museums), and sixteen digital projects. Six of the total institutions were Canadian, twenty-six were from the United Kingdom and Ireland, and thirty-two were from the United States.

Table 3

Digital Special Collection Libraries Selected for Study

Type of Institution	Library or Collection Title
University and Academic Libraries (29)	• University of British Columbia, Rare Books and Special Collections (Canada) • University of Manitoba, Dysart Memorial Collection (Canada) • University of Saskatchewan, University Archives & Special Collections (Canada) • University of Toronto, Thomas Fisher Rare Books Library (Canada) • University of Victoria, Special Collections (Canada) • Digital Bodleian (United Kingdom) • Leeds University Library (United Kingdom) • University of Cambridge, Cambridge Digital Library (United Kingdom) • University of Cambridge, Trinity College, Wren Library (United Kingdom) • University College London, Special Collections (United Kingdom) • University of Manchester, John Rylands Library (United Kingdom) • University of Warwick, Rare Books and Special Collections (United Kingdom) • Trinity College Dublin Library, Manuscripts and Archives; Early Printed Books and Special Collections (Ireland) • University of Aberdeen Special Collections Centre (United Kingdom) • University of Edinburgh, Special Collections (United Kingdom) • University of Glasgow, Special Collections (United Kingdom) • University of St. Andrews, University Library Special Collections (United Kingdom) • Harvard College, Houghton Library (United States) • Northwestern University, Schulze-Greenleaf Library (United States) • PennState University Library, The English Emblem Book Project (United States) • Southern Methodist University, Western Manuscripts at Bridwell Library to 1650/Fifteenth-Century Printed Books at Bridwell Library (United States) • University of Chicago, Early Manuscript Collection (United States) • University of Iowa, Iowa Digital Library (United States) • University of Pennsylvania, Penn Libraries, Rare Book and Manuscript Library (United States) • University of Texas at Austin, Harry Ransom Center, Medieval and Early Modern Manuscripts Collection (United States) • University of Washington, Digital Collections, Medieval and Historical Manuscripts (United States) • University of Wisconsin-Madison, Special Collections (United States)

	• Western Michigan University Libraries, Medieval Document Collection (United States) • Yale University, Beinecke Rare Book and Manuscript Library (United States)
National Libraries (6)	• British Library, Treasures and Manuscripts (United Kingdom) • National Library of Ireland (Ireland) • National Library of Scotland (United Kingdom) • National Library of Wales (United Kingdom) • Library of Congress, Rare Book & Special Collections Division (United States) • National Library of Medicine (United States)
Other Institutions (13)	• Fitzwilliam Museum (United Kingdom) • Schøyen Collection (United Kingdom) • Wellcome Library (United Kingdom) • Folger Shakespeare Library (United States) • Huntington Digital Library (United States) • J. Paul Getty Museum (United States) • Metropolitan Museum of Art (United States) • Morgan Library and Museum (United States) • New York Public Library – Rare Books Division and Manuscripts and Archives Division (United States) • Newberry Library (United States) • Schoenberg Institute for Manuscript Studies (United States) • Smithsonian Libraries, Dibner Library of the History of Science and Technology (United States) • Walters Art Museum (United States)
Digital Projects (16)	• Digital Image Archive of Medieval Music (DIAMM) (United Kingdom) • European Library (United Kingdom) • Literary Manuscripts Leeds (United Kingdom) • Manuscripts of Lichfield Cathedral (United Kingdom) • Medieval Travel Writing (United Kingdom) • Perdita Manuscripts, 1500-1700 (United Kingdom) • Irish Script on Screen (United Kingdom and Ireland) • Annotated Books Online (United States) • Botanicus Digital Library (United States) • Digital Scriptorium (United States) • Early English Books Online (EEBO) (United States) • MacKinney Collection of Medieval Medical Illustrations (United States) • Music Treasures Consortium (United States) • Pages from the Past (United States) • Parker Library on the Web (United States) • Vivarium (United States)

Note. The noted digital project Medieval and Early Modern Sources Online (MEMSO) was not included in this study, as no subscription to the project was available.

Table 3

All data was collected over a two-week window in August 2015. The use of such a small window was important to the study – libraries are constantly changing, particularly in digital environments, and for an accurate snapshot of provenance use, it is important to compare records in a small timeframe. Digital libraries are online environments, and data was collected virtually.

Although previous studies on provenance metadata used survey instruments to reach out to libraries, such methods were deemed inappropriate for this study. Surveys inform what libraries believe they are collecting in provenance metadata, which may be different or in different stages from the metadata actually available. For that reason, evaluative methods are used to assess provenance metadata across digital libraries.

In general, digital libraries have no set methods associated with evaluation (Vullo, 2010), perhaps because digital libraries are so diverse and lack consistent standards. In response, this study uses multiple methods. As the content of the metadata itself is analyzed as well as its presence, the method of content analysis is employed. Since provenance focuses on an external aspect of the book, it is bibliometric in nature. This study is therefore essentially a comparative content analysis of bibliometric materials. A cross-sectional design was used for collecting data.

Data about provenance metadata was collected for each identified library by answering the following questions:

- Was provenance metadata available through the catalog interface?
- What metadata field(s) was this information stored in?

- What extent of provenance metadata was gathered? (E.g. did it include contextual information such as dates or geographic location?)
- What metadata standard was followed?

In order to determine the metadata fields available through different digital libraries, between two and five manuscript records were compared for every digital library, depending on the number of available items for study. To fill gaps of information, "About the Project" WebPages or published articles were consulted. This helped allow for limitations, such as the fact that not all records may have provenance metadata, but the library may actively collect provenance information where available. Further, as some special collections only offer a few digital items, larger sample sizes may not be possible.

In addition to gathering data using the questions outlined above, a number of descriptive features were also noted for each collection, drawing on the correlations Overmier & Doak (1996) discovered between characteristics of libraries and the presence of provenance metadata. These descriptive features include: collection name, URL, archival affiliation, collection age (physical and digital), digital collection size, institutional classification (see above), and geographic location of the original collection. These set of factors loosely correspond to a codebook that can be applied to the data, typically used in content analysis to determine if a factor correlates with the presence of provenance. For example, certain countries or types of institutions may be more likely to provide provenance metadata than others.

Typically, content analysis faces problems because the categories used for analysis are defined by the researcher (Gorman & Clayton, 2005); however, categories used in this comparative study generally include objective information, such as

geographic location, which is less subject to bias. As a comparative study, this thesis has the advantage of being unobtrusive in its observations, and is not limited by geography.

Identifying correlations between these descriptive factors and provenance metadata are also informed by this thesis' framework. As a pragmatic study, this thesis suggests that the social and historical context of provenance in archival and special collection institutions, and the individual choices of cataloguers, influences how provenance metadata manifests today. Historical research and investigation is therefore important for this study, as a supplementary method to the overall analysis of results.

In addition to the assessment of descriptive features as identified by Overmier & Doak (1996), this study examines the correlation between certain digital features and the presence of provenance metadata in digital libraries, both because this study focuses on digital environments which allow a broader spectrum of factors to influence provenance, and because examining digital factors can help uncover if provenance metadata is being used in contemporary ways.

Digital factors selected for study were chosen based off of previous metadata research and Next Generation Catalogues (NGCs) (see Kumar, 2008). NGCs are meant to not only provide users with catalogue information, but to enhance such metadata with digital tools in order to help users understand and process knowledge. This study proposes the use of visualization to help meet many of these issues. Questions developed for the study therefore focus on visualization. When available, browsing features were employed on digital interfaces to assess NGC features. Questions used to collect data focused on digital features include:

- Were links or visualizations used to highlight metadata for users?

- What metaphors of visualization did the digital library use?
- What other pertinent information did the library have to offer? (For example, what platform is used for the collection?)

This analysis will also verify claims made in the literature review, that provenance metadata is not currently meeting user needs or expectations.

Limitations

Although a number of limitations have already been addressed over the course of research design, important limitations are re-iterated here.

Determining a representative sample of digital special collections is a difficult task, as it impossible to ascertain how many digital collections are currently in existence. Further, some collections may be difficult to classify as either a library or any other institution. Collections may also be missing from consulted lists. In order to accommodate this, the sample size of sixty-four digital libraries is comparable to other metadata studies, and is based on major recommended lists created by academic institutions and significant digital projects.

As institutions blend in digital environments, the exclusion of archives is difficult to guarantee to the fullest extent, and indeed, libraries that are joint with archives have been allowed in the study, as long as the institution self-identifies as a library as well. As hard-copy catalogues are excluded, it could be that much more provenance information is actually being captured by libraries, and is simply not reflected in the study, or that information may be in a state of backlog, and may eventually become available online with digitized items. Similarly, “invisible” provenance metadata, or encoded metadata descriptions may be lost in analysis, as this study examines descriptive standards.

Other language libraries have been excluded, so results can only apply to English-language libraries. However, English-language libraries whose manuscript and early modern materials are in other Western European languages have been included, as long as materials were neither maps nor prints. Books, manuscripts, and associated ephemera are included, as well as music, which is often found within textual manuscripts.

A notable limitation of the libraries chosen for this study is that very few academic libraries that fit the criteria are Canadian. In general, although many special collections in Canada contain medieval and early modern materials, not all libraries have digitized these collections, or have only digitized a single manuscript, which would exclude them from this study. An attempt was made to individually search as many major Canadian university libraries as possible to accommodate this limitation, and all applicable digital collections were included in the study. Australian and New Zealand libraries were excluded from the study, due to limitations of both time and sample size.

The classification of digital libraries according to type can also be difficult, as some of the digital collections created by an institutional library could be considered a digital project, or certain digital projects may be the result of multiple institutions. To resolve this issue, digital collections clearly associated with a single institution are classified as academic, whereas joint collections are deemed projects.

Only relevant questions for the evaluation of digital libraries and NGCs have been included in this study. Digital collections may offer other features related to NGCs that may appeal to users not connected to provenance. As this study focuses on medieval and early modern materials, it may be that digital collections have better provenance information for items outside these time periods, and that is not reflected in the study.

A major limitation of this study is the fact that libraries and digital environments are frequently changing. In order to constrain any effect this limitation may have had on results, data collection occurred over a strict two-week period. Although the results have many implications for libraries and the future of special collection design, it is important to remember that the information provided is only a snapshot of current metadata use.

One strength of the study lies in relational and ethical concerns, as the researcher is not directly tied to any of the institutions analyzed. Human interactions have no impact on data collection. However, determining the standards used in provenance metadata creation may be difficult, as element field names can be the same across different standards, and information on standards for particular institutions may not be available.

With the comparative study on provenance metadata outlined, it is possible to turn to the second half of the study, which builds on results from this initial analysis.

Visualization of Provenance Metadata

The second half of this study builds on the first half, by proposing visualization as a meaningful method of display that also serves to improve the current state of provenance metadata. These two halves together create an action research project.

Metadata has grown beyond simply being information in a library catalogue, through its contemporary uses in digital environments (see “Provenance as Metadata”). Using metadata in digital tools helps take advantage of modern uses of metadata. One such digital tool is visualization. Although the first half of this study seeks to establish the current state of provenance metadata in digital special collections, the second half of this study builds on this state, by proposing and testing visualization as a means of addressing limitations found in provenance metadata and meeting contemporary needs.

As discussed in "Provenance as Metadata", this thesis will focus on visualizations that use maps and timeline with some means of faceted browsing, as provenance is multi-faceted, and the axes of provenance inherently imply such metaphors. Faceted browsing is a method that allows users to access information through the use of subcategories or filters, and narrow these results accordingly (Fagan, 2010).

Research Questions and Context

A number of research questions were identified at the outset of the proposed study. This second half of the study addresses the last two questions: how might provenance metadata be displayed most productively to users on digital interfaces using visualizations; and how can visualizations help resolve current issues surrounding provenance metadata and meet user needs.

Visualizing provenance metadata presents many potential advantages to both libraries and users. As discussed in "Provenance as Metadata", visualization is a discovery tool in line with the goals of Next Generation catalogues that also meets the contemporary needs of metadata. Further, as the use of metadata in visualization necessitates some degree of standardization (Van Verchum & Pugin, 2014), it has the potential to draw out best practices in capturing provenance, thereby answering major limitations of provenance metadata in terms of lack of guidance in capturing information. In general, the literature recognizes that "visualization and provenance techniques, although rarely used in combination, may further increase […] understanding […] since [researchers] may be able to use a single tool to see and evaluate [information]" (Del Rio & da Silva, 2007, p. 732).

Although cleaning provenance metadata and inputting such information into a digital tool is a time-consuming task, it has been noted that digital projects can bring prestige and potential funding to libraries (Rimmer et al, 2008). At the same time, engaging such projects provides libraries the opportunity to address further limitations identified in provenance metadata, by creating reasons to revise current metadata and address backlog issues.

In other words, challenges associated with visualization projects may be offset by the benefits. Now is an especially ideal time to consider these projects, as special collections are already undergoing "a major cultural shift that mandates significant retraining and careful examining of priorities" (Dooley, Beckett, Cullingford, Sambrook, Sheppard, & Worrall, 2013, p. 15) with the introduction of Next Generation catalogues, Resource Description and Access (RDA) standards (Wiggins, 2012), and the advent of the Semantic Web. Further, digital libraries are increasingly introducing visualization into their metadata display (Borbinha, Pedrosa, Gil, Martins, Freire, Dobreva, & Wyttenbach, 2007; Greene, Marchionini, Plaisant, & Shneiderman, 2000; Deckelbaum, 1999; Kumar et al, 1998; Shiri, 2007). As major shifts are already occurring in libraries, it is a prudent time to consider the introduction of new tools. Visualizations are particularly apt for special collections, because of the nature of their unique materials.

In addition to enhancing user experiences, it is worth noting that visualization also has the potential to help libraries manage their metadata. Since visualizations encourage standardization of metadata, they encourage metadata to become more communicable between the four systems of metadata special collections deal with: library catalogues,

finding aids, detailed descriptions of items, and digital collection metadata (Davison, 2009).

Provenance is also apt for visualization, because of its increasing importance in digital realms and computer science (McGuinness et al, 2006). It has been suggested that "provenance research has a strong user interaction component" (McGuinness et al, 2006, p. 2), and by extension, special kinds of browsing and visualization tools should be used to highlight provenance information (McGuinness et al, 2006). Although provenance in computer science has a varied definition from book history, the axes of who, where, and when remain the same, and investigating visualization could therefore benefit both fields, especially where fields converge.

In spite of these cited benefits, little effort has been made to visualize provenance in special collections. Provenance, however, lends itself well to visualization. As explained in "Provenance as Metadata", the success of visualization depends on the use of accessible metaphors (Shiri, 2009), and provenance had at least three axes (time, place, and owners) with traditionally associated visualizations (timelines, digital maps, and networks). These visualizations that have been used individually in libraries before (see "Provenance as Metadata"), although not necessarily for provenance. Attempts to visualize provenance, however, should capitalize on combining these inherent metaphors.

There has been excitement over the prospect of combined technologies, although few combination tools exist to date. Emphasis has been particularly placed on timelines and maps, since together, these provide a rich narrative history, especially with additional multimedia overlay. The creation of interactive geotemporal tools has been celebrated by

digital humanists, as such tools "capitalize on humans' natural ability to spot patterns and relationships in visual fields" (Corby, 2010 as cited in Farman, 2010, p. 40).

Some attempts have been made to combine time and place metadata in geospatial projects within libraries. The Alexandria Digital Library, for example, created some geospatial visualizations based off of collection metadata in the late 1990s (Hill, Janée, Dolin, Frew, & Larsgaard, 1999). The Peking University Rare Book Digital Library also used temporal and spatial information elements from descriptive metadata in order to map ancient and current Chinese place names from items within the collection, through the use of digital maps (Liu, 2007).

Outside of libraries, a number of projects have been developed which engage digital mapping and timelines, such as the University of California, Los Angeles (UCLA)'s Hypercities project, which uses special collection materials in enriched mapping tools (Davison, 2009). The University of Virginia Library Scholar's Lab has also engaged such projects, as well as the Smithsonian Libraries. Geospatial ontologies are being developed for the Semantic Web (Eganhofer, 2002), and in scientific and data related fields, provenance has been visualized using Geographic Information Systems (GIS). Attempts to create scientific provenance tools such as Probe-It! have also been made (Del Rio & da Silva, 2007).

Further, a number of mapping projects related to rare books have been created: the Digitized Medieval Manuscripts Maps (DMMmaps) provides users with maps of special collection libraries around the world (Menna & de Vos, 2015), bookbinding maps from the History of Information project highlight where bookbinders were located at

certain points in history (Norman, 2015), and numerous others exist. These projects, however, do not focus on provenance, and tend to use only one visual metaphor.

An ideal visualization, however, would combine all three metaphors inherent in provenance. Combining these different visualizations has rarely if ever been done before in libraries (Johnson, 2011). With recent strides in modern technology such as the increased popularity of digital map application in humanities and social science research, and the development of easy-to-use digital timeline tools, visualizing provenance is possible now more than ever. Overall, the combination of different metaphorical visual representations is recognized as a potentially powerful tool, but as of yet, is largely underused by scholars. In order to assess the possibility of visualization to meet user needs and display provenance, numerous study methods were selected.

Study Methods and Design

The bulk of the second half of this study is a combination of an environmental scan of current visualization projects and a series of short case studies or pilot tests of sample provenance metadata in currently available digital tools.

An environmental scan attempts to identify projects of a similar nature or of significant representation within the community, in order to establish the parameters of current technology. Environmental scans are not meant to be comprehensive in scope, but instead focus on what is possible to accomplish. They serve as a stage of identification. This makes performing an environmental scan a good first step to undertake, before assessing tools for visualizing provenance.

There are literally hundreds upon hundreds of digital network, timeline, and mapping projects available online. Focus in the environmental scan was therefore placed

on projects that held similar components relevant to the study at hand. Projects evaluated must have at least two of the three axes of provenance represented (owners, time, and place), as well as a means of demonstrating relationship between these axes.

Based upon the environmental scan, which identified what features are possible in digital representation of ownership, time, and place, a set of criteria was developed for visualization of provenance metadata. These criteria are also informed by the needs of users identified in the literature,[11] and study goals. Once criteria were developed, they were used as a testing basis for case studies with sample provenance metadata on a selected set of visualization tools.

These short case studies could also be considered a series of short pilot tests, systems analyses, or as type of prototyping[12] or modeling. It is notable that visualization itself is often considered a method of analysis unto itself (Katsirikou & Skiadas, 2010), as are digital mapping methods (Powell & Connaway, 2004). In all instances, real-world sample data was used to evaluate and analyze the selected tools, so the research questions at the outset of the study may be answered. Tools are all tested and discussed individually to determine their merits and shortcomings. The testing is instrumental in nature, meaning that the phenomenon of visualized provenance is studied, rather than the individual tool on its own (Pickard, 2013).

---

[11] Criteria suggested in the literature include a tool's ability in (Chickering & Yang, 2014): federated searching, flexible interfaces which appeal to users, enriched content (e.g. images, etc.), faceted navigation (by location, date, authors, etc.), spell-check, auto-completion, recommendations, and relevancy, web 2.0 features and the potential for user contribution, and persistent links. More information on this is provided in "Visualizing Provenance Metadata".

[12] The use of digital tools and prototypes as a method is common in digital humanities disciplines, and is discussed at length in Willard McCarty's *Humanities Computing* (2005).

As the purpose of the study was to enhance digital libraries, which operate on limited budget and personnel, tools selected for study had to be open source, and easy to use.[13] Tools selected for testing were then chosen based on findings from the environmental scan. Relevant literature was also consulted to help identify important projects (Williford & Henry, 2012a). However, as projects in the digital medium are developed extremely quickly, and non-academic projects may not be covered in the literature, manual searches were also performed on the Internet for geotemporal and network visualization tools. Course syllabi, technological blogs, and other less formal content were also consulted, as well as personal correspondence with experts in the fields (such as developers at the University of Virginia Scholar's Lab). From this set, tools and projects were identified for study, and described in "Visualizing Provenance Metadata". For both the environmental scan and the choosing of pilot-test digital tools, purposive sampling is used.

The case studies are themselves somewhat heuristic in nature, although they are not usability studies. In other words, some aspects of usability are engaged for developing criteria when testing the design, but only as pertains to the representation of provenance, as opposed to overall tool experiences (Katsirikou & Skiadas, 2010). The examining of digital tools as case studies is more holistic than user studies typically are, as it focuses more broadly on effective design and representation, not only user experiences.

No predetermined sample size was selected for the environmental scan, which is explorative in nature. A few dozen projects were examined, until most features seemed to be covered, and major projects were identified. However, a sample set of four to seven

[13] Flash based tools were therefore eliminated from the study – this was done in order to accommodate both mobile functionality and screen-reader accessibility among audience members.

digital tools were selected for the collective case studies, as this number fell well within the boundaries of other evaluative studies on digital tools.

A set of sample provenance metadata also had to be prepared for testing the identified software. The selection of sample data was also accomplished through purposive sampling, as the researcher also had to prepare the data for visualization, and knew best what features were needed.

Ten manuscripts with known provenance were selected for testing provenance metadata visualization. Due to the complex nature of provenance, each of these items have multiple provenance points (see "Provenance Theory"). Ten remains a small enough number to be able to manage complex data across the end goal of four to seven tools for testing, while also being a large enough number to generate meaningful trends and conclusions from visualization – once spread across four to seven tools, the data generated is actually between forty and seventy provenance histories. The use of ten items also allows for a number of potential limitations identified in provenance metadata to be tested in visualization, such as different manuscripts that were combined together as one later in time, or gaps in knowledge of provenance.

All the manuscripts identified came from a single digital collection, the Lawrence J. Schoenberg Collection, whose physical counterparts reside at the Rare Book & Manuscript Library at the University of Pennsylvania in Philadelphia, Pennsylvania. The digital library can be browsed and searched at the Penn in Hand: Selected Manuscripts website, and includes detailed metadata information.

This particular collection was chosen as the base for sample data in part because of the vast array of codices and manuscripts available within (which range from musical

pieces to herbals and legal documents), but primarily because items in the collection had detailed and outlined provenance. Additional research was not necessary to establish this information, which would not have been possible to carry out, due to time constraints.

Choosing a singular collection from which to draw sample data is ideal for this study. The data all interacts at some point by coming into the same collection. Visualization should be able to highlight this relationship to potential users. All items were further selected by criteria set out in the framework of the study: manuscripts were of Western origin, ensured through the selection of western language items; and came from roughly the same time period of the fifteenth and sixteenth centuries.

Although the manuscripts all came with known provenance, metadata still had to be cleaned and sorted in order to be useable in visualization testing (Van Verchum & Pugin, 2014). This ensures that the metadata uses consistent formatting – for example that time periods must be represented the same way, and place names should not be abbreviated. During this process, the three axes of owners, time, and place were also isolated so that they could be input into visualization software. The result was cleaned metadata taking on the format demonstrated in *Table 4*.

Table 4

Sample Data Fields for Provenance Information

Title						
Author						
Date of Publication						
Location of Publication						
URL						
Provenance Information by Axis	*Owner 1*		*Provenance Date(s) 1*		*Provenance Location 1*	
	Owner 2		*Provenance Date(s) 2*		*Provenance Location 2*	
	Owner 3		*Provenance Date(s) 3*		*Provenance Location 3*	

Table 4

For the full sample data metadata, see *Appendix B*.

With criteria and tools identified, and sample data ready for testing, carrying out case studies was possible through inputting prepared data into selected tools. From this, a series of observations and evaluations were made by comparing results to criteria identified at the outset of the study. This analysis determined the best software for representing provenance metadata visually. Results and analyses of the case studies will be represented, as well as qualified with further observations.

If nothing currently available is capable of representing the rich relationships of provenance to place, time, and other entities, then the findings of this study can also be used to establish what digital features are possible to develop in a tool, and recommendations can be made for either improving or creating software to meet the needs of provenance metadata.

Identification and testing of tools, as well as the preceding environmental scan, were all carried out over a two-week period in August 2015. As technology and projects

advance rapidly, a small and dedicated time frame helps create an accurate snapshot of available software and prevents endless studying of new projects. All identified tools are available online, and are English-language based.

By testing the ability of identified software to visualize provenance through short case study methods, it is possible to see if visualization does enhance metadata, as suggested in the research questions. The ability of a tool to enhance metadata will be based on the ability of the tool to represent the axes of provenance, and to meet the criteria outlined by the literature and the environmental scan. Observations based on this through case study will allow the best tool to be identified and suggested for further usability testing and use.

Tools that fail to meet the conditions for study outlined above were not selected for study. Although these conditions were set in order to best meet the needs of the proposed project, it does mean that tools which require purchase or which were developed in other languages which may be better suited to user needs have not been evaluated. However, the use of open source software is important for accessibility for libraries, as most open source projects are free in both financial terms and royalties. This may provide institutions with “valued leverage over entrenched suppliers” (gbdirect, n.d., para. 29), reduce administrative and upgrade needs, and potentially increased security against viruses, data-loss, and hacks (gbdirect, n.d.), while also being more compatible with older hardware and adhering to current standards. Open source software may also provide long-term benefits to institutions, by allowing source code to be open, and avoiding issues related to the collapse of vendors (gbdirect, n.d.).

Open source software also provides benefits in the forms of robustness, as bugs are usually fixed quickly due to the availability of the source code; providing choice to institutions, as pressures to change or upgrade software according to vendors can be lessened; auditability, as the source code is available to users, so claims about security, adherence to standards, and adaptability of the tool can be verified independently of the tool creators; flexibility, as institutions can modify software to suit their own needs, the tool is less likely to be dependent on other related products, and institutions are not tied to a single vendor; and support, as institutions may potentially seek support from other sources than the author of the tool itself (gbdirect, n.d.).

The use of these methods reinforces the overall pragmatic methodology of this study, as the evaluation and proposal of a visual prototype based on library provenance metadata is a real-world oriented practice (Mackenzie & Knipe, 2006). This second half of the study also uses grounded theory, as is common in case studies. This means that both data collection and analyses are ongoing at the same time. This method helps bring theories of provenance together, and will hopefully lead to the creation of a recommended model to capture and display provenance metadata.

Limitations

This section highlights limitations of the second half of this study, many of which have already been addressed. Examining them together, however, ensures transparency in research design.

In any visualization carried out on library metadata, a major challenge is cleaning the metadata so that it can be used to produce the visualization (Van Verchum & Pugin, 2014). Provenance metadata, however, provides another challenge as a field potentially

missing information and needing intense research (see “Provenance Theory”). In this study, data with known provenance was selected to circumvent this problem. However, this makes provenance visualization a potentially expensive and time-consuming undertaking for libraries at large, unless such projects bring funding and prestige.

Although visualizations may provide benefits such as interoperability and enhanced user experiences, as this project seeks to investigate, it may be that provenance metadata in its current state is too difficult to visualize with any meaning. This can only be determined through testing, however.

Selecting a sample size is incredibly difficult for such a study, as it is unknown how many tools are available for testing, and new digital projects are created daily. Performing an environmental scan, which is investigative in nature, helps counter this problem by allowing the researcher to explore relevant projects and tools. The selected number of four to seven tools for evaluation is within the spectrum of comparable studies.

It is possible that important projects may be missed within the outset criteria, particularly as environmental scans can be quite general in nature. Similarly, researcher bias is a potential limitation of this project, as multiple steps in the process require the researcher to make an informed decision. The use of several sources to identify tools and projects helps counteract this limitation.

Outside of missing projects within the research framework, the criteria set out for selecting tools could theoretically also lead to the potential exclusion of tools that may be more robust and better suited to provenance representation. Due to ease of use and accessibility, important values to libraries, however, these tools are likely not viable if they do not fit the criteria set out by the research design.

This project is also limited by time – it may be that tools are currently being developed and may be available for use quite soon, but due to time constrictions were not evaluated in this project. This time frame is necessary, however, in order to assure an accurate set of results.

With a complete research plan in place, it is possible to turn towards "Provenance Metadata in Digital Special Collections", and "Visualizing Provenance Metadata" which presents the results and analyses of the entire action study. Examining the current state of provenance metadata in digital special collections allows us to determine if provenance metadata is being used to its full potential, and if visualization can truly enhance provenance information.

CHAPTER FIVE: PROVENANCE METADATA IN DIGITAL SPECIAL COLLECTIONS

In order to answer the questions posed at the outset of this thesis, a two-part action research study was devised (see "Methodology"). The results of this study can help establish the current state of provenance metadata in digital special collections, and propose new ways of enhancing that information for prospective users. The results and analysis of the first part of the study, which focuses on the use of provenance metadata with digital special collections, are presented here. Data was gathered and analyzed using the methodology outlined in "Methodology". Discussions based on findings are also presented. The second half of the study is explored in "Visualizing Provenance Metadata".

Data discussed in this chapter was gathered over the course of two weeks from August 13, 2015 through August 27, 2015. This data was sorted and analyzed in three stages. First, information solely related to provenance information of each selected collection was analyzed. This information included: the presence of provenance metadata, the metadata elements used to store provenance information, the axes offered in provenance information, and the metadata standard used by the collection.

Following this general analysis of provenance metadata in digital special collections, the presence and available axes of that provenance metadata was compared to features found within those collections. These features were partially identified from previous analyses of provenance metadata in special collections (see Overmier & Doak, 1996), the historical investigation of provenance theory within special collections (see "Provenance Theory"), and contemporary understandings of metadata (see "Provenance

as Metadata"). Features were then divided into two sets: descriptive features and digital features of the collection.

Descriptive factors included:

- any obvious archival association within the collection,
- the physical and digital date the collection was established,
- the number of digitized manuscripts and/or rare books within the collection,
- the type of institution the digital collection came from, and
- the geographic location of the host collection.

Digital features included:

- the obvious use of digital platforms to present collections,
- the presence of links within metadata,
- the presence of visualization within metadata, and
- the metaphor used for visualizing metadata (where applicable).

In order to best understand the use of provenance metadata within digital special collections as well as descriptive and digital features and their relationships to that metadata, an exploratory analysis was carried out. Each of the factors listed above was treated as a variable, and explored in a simple statistical analysis of collections based on the presence of provenance and the axes of provenance available within that metadata. In other words, data was analyzed in order to ascertain how individual factors may or may not statistically correlate with the availability and richness of provenance information. Information was summarized in a number of tables, and then displayed in qualitative bar charts for better study (see *Appendix C*).

It is of note that very complex analyses are possible with the data collection, particularly of a multivariate nature where different factors are cross-analyzed. For example, the factor of different types of institutions could be analyzed against the factor of different geographic locations in the context of provenance metadata. Perhaps academic institutions in the United States capture provenance more often than academic institution in the United Kingdom and Ireland. Although such analyses are possible, these complex cross-interactions extend beyond the scope of this thesis. Instead, only individual factors and their relationship with provenance metadata will be considered. Results of this study are examined below.

Provenance Metadata

Data revealed that a vast majority of digital special collections display provenance metadata – a full 80% or fifty-two collections of the sixty-five studied (see *Figure 5*).

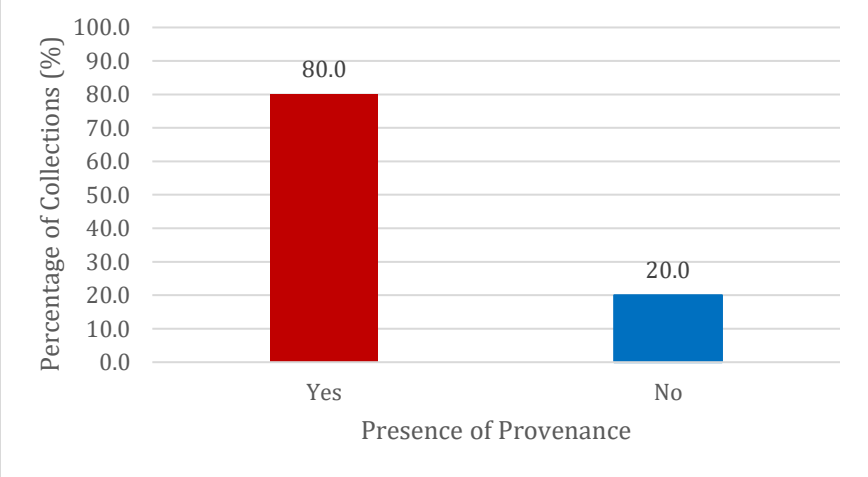

Figure 5: Bar graph showing the presence of provenance metadata in digital special collections

These numbers support the suggestions made within the literature, that the presence of provenance metadata has improved over the past twenty years (Pearson, 2005). Overmier & Doak (1996) found within their study that only 51% of special collections maintained provenance records online. In just under twenty years, the presence of provenance

metadata has increased almost 30% in online environments. This finding also supports the assertion that provenance theory has enjoyed an increased focus and importance within book history disciplines, and consequently libraries, over the same time period (Pearson, 2005).

Axes within Provenance Metadata

In addition to an increased presence of provenance, the literature suggests that the quality of provenance metadata has also improved over time (Pearson, 2005). If this is true, and the quality of provenance metadata has improved, this further suggests that the digital tools of Next Generation Catalogues (NGCs) are at an apt time to launch.

In order to assess the quality of provenance metadata, provenance metadata was analyzed to determine which of the three axes of provenance: name, location, and date (see "Provenance Theory"), are represented in each digital collection.

As information was collected and sorted, five different categories were discovered for provenance metadata, based on those three axes or some combination thereof. Each category was assigned a code, for ease of reference (see *Table 5*). According to this system, the higher the number within the code, the more axes were present in the provenance metadata of the digital collection.

Table 5

Key for Codes Assigned to Provenance Metadata based on the Axes present in Digital Special Collections

Assigned Code	Axes represented
1A	Owner name only
1B	Location only
2A	Owners and Dates only
2B	Location and Dates only
3	Owners, Locations, and Dates

Table 5

Each category includes incidents where *only* those axes are present. For example, category 2A includes provenance metadata that displays owners and dates, not owners only, dates only, and owners and dates. These codes were maintained throughout the rest of data analysis to ensure clarity of information.

The axes available in provenance metadata do suggest some measure of quality of provenance information. If more axes are present, it is likely that more research was conducted and provided in the provenance metadata. However, this relation is not always perfect. For example, it is possible that even institutions that offer the three axes of provenance within their metadata may have only offered some information, or left out data. For example, early provenance information which may require more time to research may have been omitted. Further, information could be clouded – an owner may have lived in several locations through his or lifetime. If a location is given, does this refer to the place where the item next changed hands? Or the place the item spent the most time? Should only countries or continents be marked in these cases? Similar questions could apply to dates.

However, the assumption of more axes offered suggesting better quality metadata is the best gauge of quality this study could offer within the designated time restraints – assessing quality of provenance metadata on more nuanced levels would require intensive research which was not feasible within study limitations. Collections were sorted according to which code their provenance metadata fell in, and the statistical results were illustrated in a bar graph.[14]

[14] Note that axes themselves were not visualized in these graphs, but rather the percentage of collections that provide metadata with axes of the corresponding code. In consequential graphs, histograms will be used to visualize information, as each code is part of a continuous data set evaluated against selected variables.

Interestingly, just under half of collections offered three axes of provenance metadata (41.8% or twenty-four collections) (see *Figure 6*). Provenance metadata including only the names of owners was the next most common (27.3% or fifteen collections), followed closely by the combination of name and dates (21.8% or twelve collections). Collections offering provenance information in the form of locations or locations and dates were significantly few in comparison (only 3.6% or two collections). It is likely that the concept of ownership is dominant over location in many special collection interpretations of provenance.

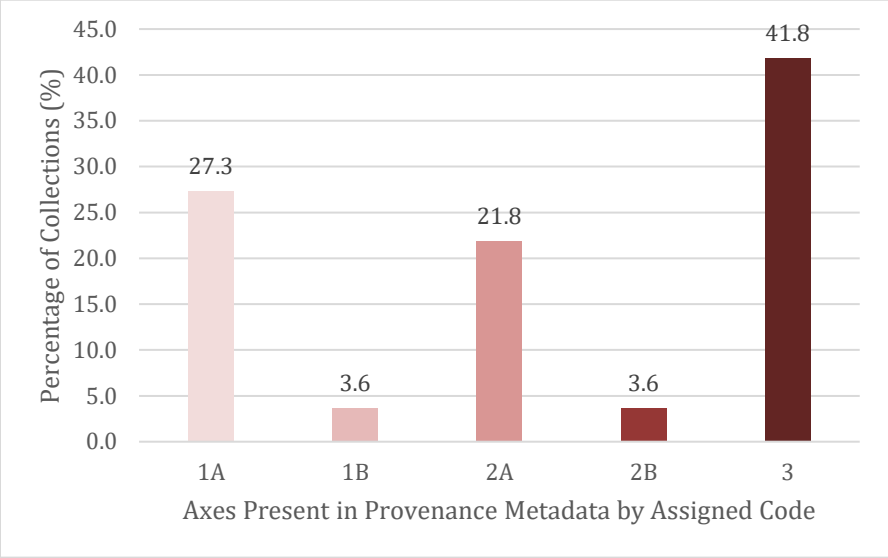

Figure 6: Bar graph showing the axes offered in provenance metadata within digital special collections

This data is revealing – since categories 1A and 2A are so similar in nature and may indeed have cross-over (dates may only refer to the lifespan of the author, for example), it supports the conjecture within the literature that often provenance metadata is only relegated to owner names (Pearson, 1994; Lundy, 2008). Together, these two categories account for 50% of all collections. In contrast, emphasis on contextual location information yields very little emphasis, accounting for only a total of 7.2% of all collections. However, the fact that the use of three axes trails behind so closely the

majority of collections at 41.8% implies that metadata being captured is slowly improving and may surpass less nuanced metadata.

This suggests that now more than ever is an apt time to visualize provenance metadata. It is more present than ever in digital collections, with more axes and therefore potentially increased quality.

Provenance Metadata Elements

Twenty-five metadata elements were identified within the fifty-two digital special collections for capturing provenance metadata.[15] These elements were:

- Added Author
- Administrative history
- Associated Name (Former Owner)
- Author/Owner
- Binding Details
- Comments and Notes - Provenance
- Contributors
- Description
- Description - Provenance
- Donor
- Former Owners
- General Note
- History - Provenance
- Inscriptions
- Names
- Notes
- Notes - Ownership notes
- Other name
- Ownership
- Ownership - Provenance
- Ownership History
- Places
- Provenance
- Related Names
- Within a Paragraph

Although a great diversity of elements exists with digital collections, many of these elements were actually only used once throughout the entire study. Several collections also made use of more than one metadata element in their capture of provenance information. These factors help explain such a large variety of elements.

The most frequently used metadata element in digital special collections to capture provenance was aptly named "provenance" (see *Figure 7*). Almost a third of

[15] As markup and embedded metadata were not included in this study (see "Methodology"), these elements refer to immediately visible labels within digital collections.

provenance metadata was captured with this element (32.9% or twenty-three instances of use), suggesting that close to half of collections use it (44% or twenty-three collections). This is significant for an element that has no firm tie to any metadata standard – it seems almost half of collections have adopted their own practices for capturing provenance metadata, and recognized a need to provide such information.

The next most used elements were free-text forms within paragraph descriptions somewhere in the collection (14.3% or ten instances), "description" (11.4% or eight instances), and "note(s)" (10% or seven instances). The complexities of provenance lend it well to free-form paragraphs and similarly unstructured elements such as "description" or "note(s)", likely explaining the popularity of such fields.

The prevalent use of "description" and "note" elements and variations of those elements (such as "Notes – Ownership history") support suggestions within the literature that a great deal of provenance information is simply relegated to general notes (Lundy, 2008), making the information more difficult to locate and far less structured.

The use of so many elements overall suggests that provenance metadata is still largely unguided for digital special collections, especially as digital libraries themselves have so few standards. Element names varied greatly, from "Added Author", "Inscriptions", "Ownership History" and "Binding Details". These different names suggest that different collections may place emphasis on different aspects of provenance – some are more interested in provenance evidence than provenance information. Others might see donor information as more important than general ownership. Both location and owner axes feature in metadata element names, however.

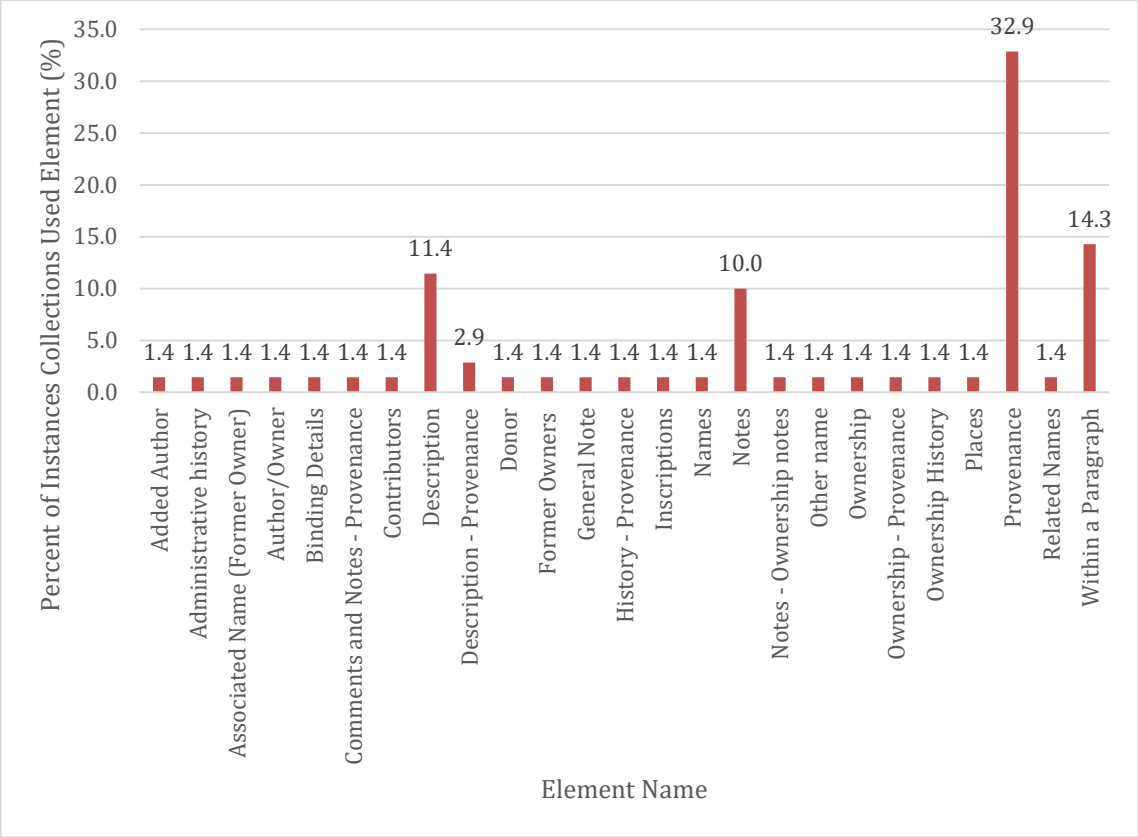

Figure 7: Bar graph showing the popularity of metadata elements used to capture provenance within digital special collections

In general, it seems that further guidance would benefit provenance metadata element usage, although a trend seems to be emerging to move away from general "description" and "note(s)" towards "provenance" itself as an authoritative field, although this field still remains largely undefined.

Use of Provenance Metadata Standards

At a glance, it is difficult to determine from the metadata elements used to capture provenance which metadata standards are actually being used. This is especially true if the institution did not offer further information on their cataloguing and metadata

practices elsewhere on their website.[16] An attempt was made to determine metadata standards used within digital special collections, regardless.

Almost all standards showed a strong presence of provenance metadata.[17] In line with the metadata elements used, data collected revealed that the majority of digital special collections used their own set of local or mixed elements (40.2% or thirty-five of those collections offering provenance) (see *Figure 8*). This finding is to be expected. Digital collections have no fixed standards, and many digital projects with self-generating metadata often create their own set of standards, in order to meet project and user needs.

The second most frequently used metadata standard is the Anglo-American Cataloguing Rules, Second Edition (AACR2) and/or the Resource Description and Access (RDA) (12.6% or eleven instances of those collections offering provenance). The high frequency of AACR2/RDA use is understandable, considering their prevalence within library cataloguing systems. Digitized collections are likely to use catalogue metadata to save time and money. A high presence of AACR2 / RDA also helps explain the high frequency of "note" metadata elements, and provenance emphasizing the axis of ownership, as these are features of the standard for capturing provenance metadata.

The third most frequently used method of displaying provenance information was through descriptions within essays (10.3% or nine instances of use). Interestingly, every essay description found in the data provided provenance information. This is likely

---

[16] In instances where extra metadata elements seemed to be added outside of a predominant standard, the collection was classified as using mixed and/or local elements.

[17] Notably, the National Library of Medicine (NML) Metadata standard was the only standard never used to capture provenance information (although only one instance of use for NML was found overall). This is likely because provenance is not a focus or priority for most materials captured by NML, which was actually "designed for use with electronic resources published by the NML" (AIMS, 2014, para. 1). Consideration of rare books was not likely in creating the standard.

because essays provide an ideal method of explaining the complexities and nuances of provenance information, as any format or length of explanation can be used.

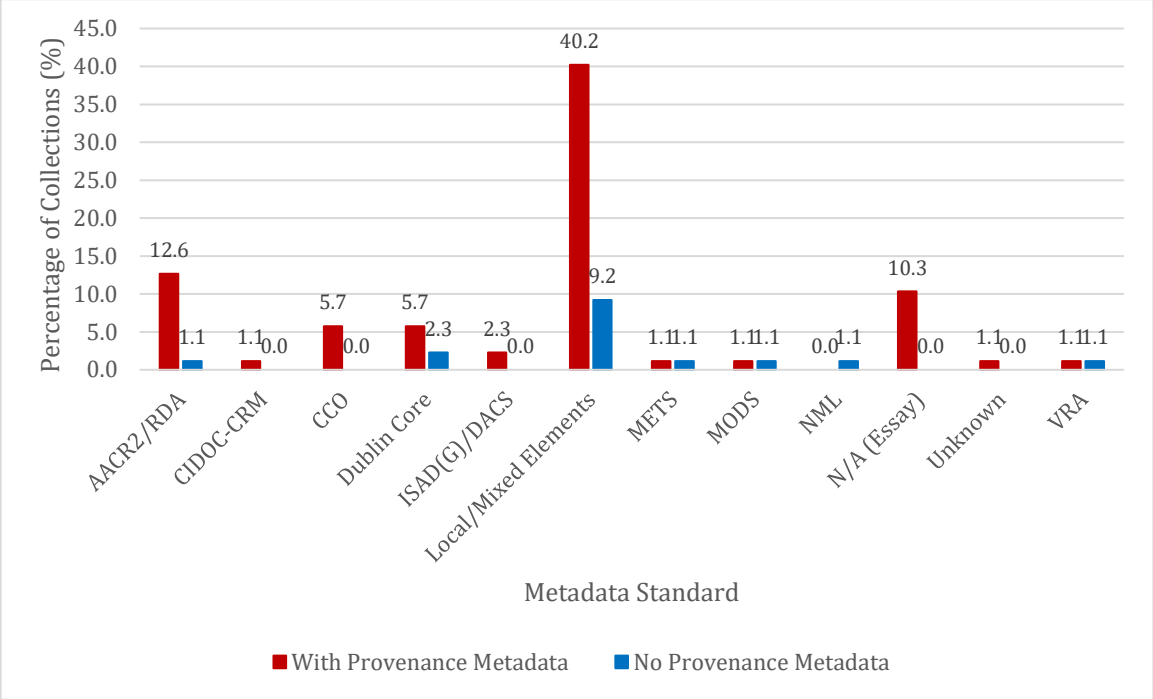

Figure 8: Bar graph showing metadata standards used to capture provenance within digital special collections

Both archival based standards (ISAD(G)/DACS) and museum based standards (CIDOC-CRM and CCO) always included provenance within their metadata. Considering the importance of provenance in both these disciplines (see "Provenance Metadata"), these findings are expected.

Interestingly, very few institutions made use of archival-based standards, considering the nature of special materials – there was only 2.3% frequency of use for ISAD(G)/DACS (or two institutions). Museum-based standards were actually more prevalent, with 1.1% frequency of use for CIDOC-CRM (or one institution), and 5.7% frequency of use for CCO (or five institutions). It may be that museum libraries have made larger strides in digitizing their materials at this point in time than archival ones.

Metadata standards with more digital applications, such as Dublin Core, METS, MODS, and VRA were more evenly split in their use of provenance metadata, compared to more traditional standards. As standards with a larger emphasis on digital environments, it is understandable that provenance of rare books and manuscripts has been largely overlooked. The fact that provenance metadata is available half of the time these standards are used may point towards a dawning recognition of the importance of provenance even in digital mediums, as discussed in "Provenance Theory".

The large number of mixed and local element collections can be broken down into more specific mixes of standards. This data reveals that collections have drawn influences from all kinds of different sources and standards, and a strong preference towards local interpretations or creation of elements still exists (18.4% frequency of use of local elements within collections, or eleven collections).

What is surprising about this finding is that almost a third of institutions using local elements do not include provenance metadata (5.7% frequency of use out of all standards, or five collections). The literature suggests that local elements are commonly used (Lundy, 2008), which was true. The literature also suggests that provenance was found most commonly where local elements are (Lundy, 2008; Pearson, 1994), which was not necessarily true. Close to all institutions using mixed elements offered provenance, as opposed to only two-thirds (or eleven collections) of local elements. This suggests that guidelines within standards have indeed improved since Overmier & Doak's 1996 study.

Axes within Provenance Metadata and Metadata Standards

With some sense of the standards used, the relation between standards and the axes provided within provenance metadata was briefly explored (see *Figure 9*). Certain standards may consistently use more axes in their provenance metadata.

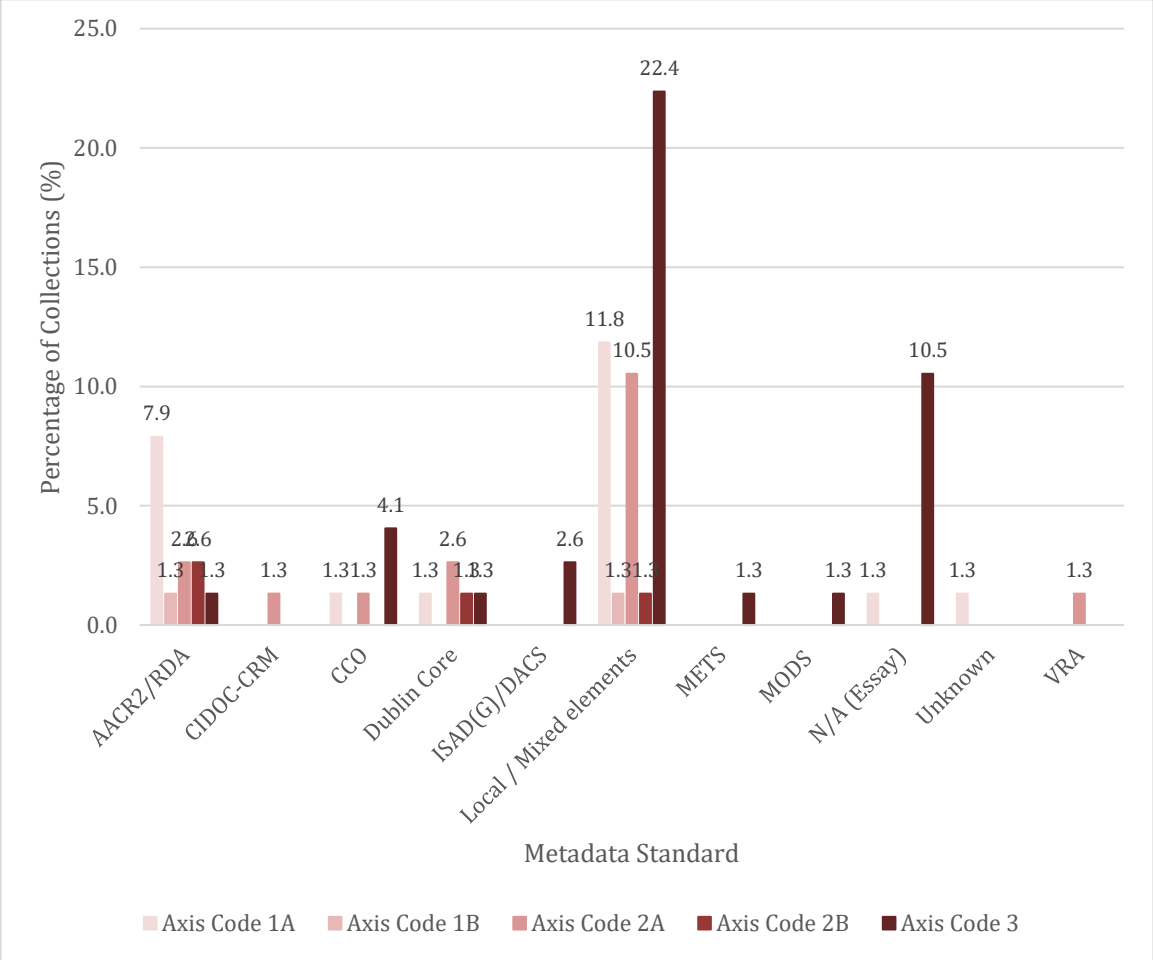

Figure 9: Bar graph showing the axes within provenance metadata by standard

Interesting patterns emerge when examining the available axes within provenance metadata are compared by standard. Local or mixed elements had the most variety of available axes, with figures strongly echoing overall study findings related to the distribution of axes within provenance metadata. The majority of collections using local and mixed elements (22.4% or eighteen institutions) used all three axes of provenance,

closely followed by focus on the singular axis of owners (11.8% or nine institutions) or the dual axes of owners and dates (10.5% or eight institutions).

When examining the use of AACR2 / RDA, however, a noticeable emphasis is placed on the axis of owners. Of the twelve collections using AACR2 / RDA, half of them only provided owner information. This supports suggestions in the literature, that many libraries only provide owner names, and no additional information (Pearson, 1994; Lundy, 2008). In fact, only one institution using AACR2 / RDA provided all three axes of provenance, suggesting guidance on capturing provenance information within these standards is still lacking.

In contrast, those institutions which did not follow a metadata schema, but instead only offered free-form essays with information on items or collections almost all provided three axes in their provenance information (eight institutions of nine). The format of an essay likely lends itself well to the inclusion of detailed provenance information, accounting for this increased number of axes.

As expected, those collections using archival standards (ISAD(G)/DACS) all provided three axes of provenance. The emphasis of provenance in archival practices makes all aspects of provenance important. In museum standards, the single institution using CIDOC-CRM provided only owner and date information. Of those institutions using CCO, the majority provided three axes of information. Museum standards also tended towards providing more axes of provenance in their metadata. Overall, institutions using digital standards showed a wide variety of availability of axes within provenance information, emphasizing the lack of guidelines within digital standards.

With a better understanding of the state of provenance metadata within digital special collections, it is possible to examine what other factors might have a correlation with provenance metadata.

Descriptive Factors and Provenance Metadata

Key findings of the relationship between descriptive factors and provenance metadata are discussed in this section. *Appendix C* provides graphs of the data. Most data suggested light trends in terms of the use of provenance metadata when different factors were considered, instead of fixed or determined rules.

Archival Ties

Archival theory is intrinsically tied to provenance. Further, the development of archives may have influenced special collections, due to the nature of unique items in both institutions (see "Provenance Theory"). For this reason, examining the association of a digital collection with archives may reveal some influence on the presence of provenance metadata, and what axes of provenance are available within that metadata. Increased presence and numbers of axes are expected in collections with explicit associations with archives.

Most collections examined did not have expressly visible ties to archives as defined by the study (see "Methodology"), but those that did almost all provided provenance metadata (93.4% of collections with archival ties). Further, that metadata tended to provide all three axes of provenance (50.0% of collections with archival ties and provenance metadata). In cases where not all three axes were used, there was a sizeable chance that at least two axes were present (see *Figure C1*). Findings support the

theory that collections with archival ties tend to provide provenance metadata, and that that metadata is more comprehensive.

Physical Collection Age

The second factor examined against the presence of provenance and available axes within that provenance metadata was the date the collection was established. As this study examined digital special collections, that factor could be interpreted two different ways: the date the physical collection was established, or the date the digital collection began. Both of these factors matter. Digital collections are based on physical collections. Physical collections may have pre-established metadata, which is simply translated to the digital medium. The age of the physical collection is therefore relevant. On the other hand, the date when the digital collection was created may also affect portrayal of metadata, as digital standards have changed over years, as has the prominence of provenance theory in digital mediums Each of these was examined in turn where data was available.

Overmier & Doak's study in 1996 found the older the institution, the more likely it was to provide provenance in print catalogues. This was theorized to be because older institutions had more time to both establish the provenance of their collections, and create metadata based upon their research. Interestingly, however, findings from this study suggested the opposite (see *Figure C2*). The newer the physical institution, the more likely it was to provide provenance metadata over older counterparts. In fact, all institutions under 50 years of age provided provenance metadata. Further, younger institutions were also more likely to provide metadata containing all three axes of provenance (50.4% of institutions under 50 years of age).

The discrepancy of these findings might be accounted for in this study due to the nature of the digital medium. Older institutions may still keep their provenance metadata in print format due to backlog that is not available online, accounting for lower presence of provenance and number of axes provided within that provenance metadata online. Newer collections, on the other hand, have less backlog and may tend to adhere to newer understandings of provenance theory. This makes them more likely to provide provenance metadata.

In further support of this reasoning, only institutions under 100 years old included contextual information in the form of location in their provenance metadata. This finding aligns itself with the development of provenance theory, which has only expanded to include contextual information in the past century (see "Provenance Theory").

Institutions over 100 years old were most likely to overlook provenance metadata (19.2% of collections over 100 years old), although when it was provided, it tended between three axes (38.4% of collections over 100 years old) or only a singular axis (42.3% of collections over 100 years old). It may be that older institutions used older guidelines for provenance metadata that suggested only one axis of provenance information, but are slowly moving towards more complete provenance metadata in line with current interpretations of provenance, accounting for the presence of more axes of provenance.

Digital Collection Age

When examining the parallel factor of digital age of a collection, similar results to physical collection age were found (see *Figure C3*). The younger the digital age, the more likely the collection was to offer provenance metadata with three axes (83.1% of

collections under 10 years offered provenance, and 44.6% of those collections had three axes of provenance). As with newer physical collections, if available metadata did not have three axes in newer digital collections, it only had one. This suggests new collections have a tendency to embrace newer understandings of provenance, or in lieu of that, only provide basic information.

Older digital collections were least likely to offer provenance metadata (75.7% of collections over 10 years old), or context in their provenance metadata, such as location-based information. This suggests older digital collections are also less likely to take advantage of newer theories of provenance. Perhaps their systems are less able to accommodate such nuanced information.

Collection Size

Overmier & Doak's (1996) original study found that larger collections were more likely to provide provenance metadata. This was thought to be because larger institutions were more established and potentially had more resources to capture provenance information. The number of digitized manuscripts and/or rare books in a collection was therefore compared to the presence of provenance metadata and the axes available within that metadata in this study.

In another contrast with Overmier & Doak's (1996) study, however, findings from this project suggested that the smaller the collection, the more likely it was to provide provenance metadata (86.2% in collections under 100 items) (see *Figure C4*). Further, smaller collections also tended to provide metadata containing all three axes of provenance. It is likely easier for smaller collections to amass provenance metadata for

their items, as there are fewer of them. Larger collections may keep provenance in backlogs.

Larger collections, however, were also more likely to offer contextual location information than their smaller counterparts. This may be for the reasons Overmier & Doak (1996) suggest - larger collections tend to be more well established with more resources, allowing them to interact more with changing interpretations of provenance.

Type of Institution

Institutions were identified according to four different types, as established in previous evaluative studies on digital libraries (Chowdury & Chowdury, 2000). These types of institutions were academic-based digital collections, national and governmental digital collections, independent or other collections (in this study, largely museums), and digital collections born of academic and other scholarly digital projects. Findings suggested that the type of institution a digital collection is based on does have some bearing on the presence of provenance metadata, and the axes available within that metadata (see *Figure C5*).

Museum and other collections were the most likely to provide provenance metadata; however, this metadata did not necessarily offer multiple axes of provenance. Further, contextual information such as location was hardly ever offered. Academic institutions were the second most likely to offer provenance metadata, but in contrast with other collections, that metadata was likely to contain all three axes of provenance. Contextual information such as location was also significantly featured where fewer axes were present. It seems that academic institutions pay close attention to modern definitions

of provenance, and consequently place emphasis on providing multiple axes of provenance.

Digital projects had a high chance of providing no provenance. Where provenance metadata was offered, it split between containing three axes, or only a single axis. These findings make sense, as digital projects cater to specific purposes and audiences. These may not focus on book history. Digital projects are also more likely to use software developed by third-party companies. These companies may have limited experience with library-based metadata, leading to the omission of fields such as provenance.

National libraries were the least likely to offer provenance metadata, also likely due to their specialized nature, but they did tend to offer provenance information with two or three axes. These findings make sense considering the development of national libraries as often strictly separate entities from national archives.

Geographic Location

The final descriptive factor examined with the presence of provenance metadata and the number of axes of provenance available within that metadata was the geographic location of the host collection. Institutions were sorted by country: Canada, the United States (US), and the United Kingdom (UK) and Ireland (see *Figure C6*).

All Canadian institutions provided provenance, although very few institutions were featured in this category. Institutions were fairly evenly distributed in terms of the axes provided within that provenance metadata. These findings suggest that the provenance renaissance Canada experienced within archival institutions also affected special collections (see “Provenance Theory”), leading to increased presence of provenance metadata.

Collections from the United Kingdom and Ireland were the second most likely to provide provenance metadata. Generally, that metadata contained at least three axes of provenance. Collections from the United Kingdom and Ireland therefore seem very responsive to modern provenance theories. This may be in part because of the wealth of rare books and manuscripts found there.

Collections from the United States were the most lacking in provenance metadata, and also the least likely to offer provenance metadata with more than one axis. Historically, special collections in the United States take after archives (see "Provenance Theory"), so this finding is interesting. It may be the large sample size of institutions from the United States has affected this outcome, but this finding may also indicate that modern provenance theory is slower to be embraced in digital collections from the United States.

Digital Features and Provenance Metadata

The relationships between digital features such as digital platforms, linked metadata, and metadata visualizations may shed light on contemporary uses of provenance metadata. This section explores the relationship between these features and the presence of provenance metadata and axes available within that metadata.

Digital Platforms

Some collections included in the study made use of notable third-party digital platforms. Digital platforms are software or hardware used to create a website. Some of these come with pre-determined means of displaying digital collections, including how and what to display in metadata, which can affect the presence of provenance

information. Three categories were identified in the data: collections using ContentDM, collections using LUNA, and collections using other platforms (see *Figure C7*).

ContentDM is a digital collection management system (DCC, 2015), and a Drupal[18] resource (Drupal, n.d.). A fee is required to use it. ContentDM offers some guidelines on what metadata to use in collections. It was the most common digital platform (62.5% of collections using a digital platform chose ContentDM), although half of the collections using ContentDM offered no provenance metadata. Where provenance metadata was offered, it generally offered two axes of provenance metadata.

LUNA was the second most commonly used digital platform (25.0% of collections using a digital platform chose LUNA). LUNA is a tool used to manage primarily image-based media. Collections using LUNA can link multiple records to an item, or create their own metadata schemas (LUNA Library, n.d.). Results of this study found that all collections that used LUNA also provided provenance metadata. This is of interest, since LUNA allows collections to create their own standards (LUNA, n.d.). Further investigation reveals provenance metadata in LUNA platforms was split between one or three axes. This suggests that LUNA encourages users to provide provenance information. The result is certain collections offer scant provenance information, while others become more invested, and offer richer information through more axes of provenance.

Examples of other digital platforms used by collections include VuFind and WebGenDB. Collections in this category only provided provenance half of the time. Like

[18] Drupal is an open source software available to individual users or institutions. It is used to help create and manage many different kinds of web pages, and is designed to be usable even for those without strong technical skills.

collections using ContentDM, provenance presence and the axes available in that provenance metadata within this category seems to rely on the discretion of the collection.

Linked Metadata

Another important digital factor examined in this study is the presence of links within metadata. If collections with linked data also provide provenance, then introducing other digital tools is very feasible, as the data already has built-in connections.

Results revealed that half of collections with provenance metadata also had linked metadata (see *Figure C8*). No corresponding link seemed to exist between lack of linked metadata and lack of provenance metadata. Provenance metadata was generally split between three axes or one axis. Interestingly, however, strong emphasis was placed in collections with both linked metadata and provenance metadata on location-based contextual information in the provenance metadata (9.1% of collections with linked metadata). This suggests that linked metadata is present in collections with richer metadata or who have embraced more modern theories of provenance.

Expanding on this digital factor, the amount of linked metadata used in collection records was also investigated (see *Figure C9*). Collections were classified according to three categories: those with scant linked metadata (only one field used links), those with moderate linked metadata (two or more fields used links), and those with heavy linked metadata (where all or almost all fields used links).

Most collections with linked metadata fell into the scant category (40% of collections with linked metadata), but these collections were split between providing three axes of provenance or a single axis. Collections were second most likely to fall into

the moderate category (32.0% of collections with linked metadata). Similarly, collections were also split between three and one axis. Few collections fell into the heavy category (28.0% of collections with linked metadata), although those that did were split fairly evenly in the number of axes provided within the provenance metadata. In general, the less linked metadata there was, the more likely provenance metadata had two or three axes. Overall, these findings suggest both that richer collections are more likely to offer linked metadata, but also that it is easier to create linked metadata with three axes of provenance when there is less overall data to manage.

Metadata Visualization

The next factor examined was the presence of metadata visualizations (see *Figure C10*). Visualization is one of the most advantageous uses of metadata (see "Provenance as Metadata"). Very few collections offered metadata visualizations (13.6% of collections), but where they were available, provenance was always present. The majority of these collections offered at least three axes of provenance, and almost the same number offered two axes. This suggests that where provenance metadata is available with visualizations, it corresponds with metadata that provides more axes of provenance. Emphasis was also placed in the data on location-based metadata in provenance where visualizations were present. This means that collections using visualization also embraced more modern ideas of provenance theory. Contemporary understandings of metadata go hand-in-hand with contemporary understandings of provenance.

Visualization Metaphors

Certain metaphors were identified in the visualizations used. These were classified in categories created based on the axes of provenance: timelines, based on the

axis of time; digital maps, based on the axis of place; networks, based on the axis of ownership; any mix of the above metaphors; and any other metaphor used in the visualization (see *Figure C11*).

Timelines were the most used metaphor by collections with visualizations (30.8% of collections with visualizations). Collections were fairly evenly split according to the number of axes of provenance provided within the metadata. Digital maps were the second most likely metaphor to be used (23.1% of collections with visualizations). These tended to be used by collections offering only one axis of provenance, unsurprisingly, with focus on location-based information. Networks were the most uncommon visualization (15.4% of collections with visualizations), although these collections tended to use at least two axes in provenance metadata. Interestingly, no collection combined metaphors in any way. This suggests that the blending of metaphors for complex metadata is still an untapped tool in digital special collections.

Other metaphors found in this study included sliders, or visual projects accessible outside of the digital collection. These tended to be used where provenance metadata provided more axes, unlike timelines and digital maps. It may be that timelines and digital maps are not adequate to portray complex information on their own. These findings also support the literature in suggesting that visualizations, more generally, are a useful tool for making sense of complex data (Chen et al, 2014).

This data supports that both timelines and digital maps are established provenance metaphors in visualization. In both cases, visualizations were used with different purposes – either browsing, or highlighting information about an item, but never both.[19]

[19] To learn more about how these visualizations were used in their respective collections, see "Visualizing Provenance Metadata".

Further, these metaphors are rarely used with complex metadata, suggesting that the metaphors are not ideal for supporting complex ideas. It is possible that mixing metaphors may help alleviate this problem, as more complex visualizations are possibly better suited for complex information. There has been little mixing of metaphors to date, however, leaving provenance as a concept with multiple axes largely untapped in digital mediums. Finally, some visualizations are only available outside of collections, making those visualizations less useful to users.

Further Discussion

In light of the findings presented in this study, a discussion of major challenges and recommendations made in the literature for recording provenance as metadata follows here. This discussion can help inform findings of this study, and highlight needs of both the library and book history communities. These needs, in turn, inform any digital tools sought to be brought into NGCs for provenance.

Challenges of Provenance as Metadata

In "Provenance as Metadata", a number of issues were identified in the capture and cataloguing of provenance information. These include choosing not to provide provenance metadata for various reasons, and difficulties in selecting from a large number of different standards, attempting to combine these standards, or opting not to use standards at all. Cataloguing provenance can also be difficult due to unclear guidelines within standards, leading cataloguers to be unsure of the depth of information to capture, and conflicting community priorities, leading to certain axes of provenance being ignored. Lack of time and funding for in-depth cataloguing, as well as backlog issues, also present challenges.

It is useful to revisit these challenges in light of the results of this study. New issues also come to light, such as problems with provenance as metadata due to the use of embedded metadata, resulting in provenance information not being visible to users.

Why is it some items or collections may not have provenance metadata at all? The most obvious reason for omitting provenance is, of course, the difficulties inherent in assessing provenance (see "Provenance Theory"). Incomplete data or data that requires large amounts of time and research to assess may result in a collection deciding to just omit provenance. Similarly, even collections which provide provenance metadata may not include it for items where such information is simply not known. This may be the case for the 20% of collections in this study which provided no provenance metadata.

Another factor which any affect the presence of provenance metadata is the complex relationship and history of special collections to archival institutions. Public libraries almost never capture provenance metadata, due to their inherent nature. Public libraries focus on circulation, not collecting unique items. They may also choose not to capture provenance information because such metadata would only be relevant to a small group of users (Bruni, 2011). On the other hand, archival institutions, which focus on provenance as a core fundamental to archival theory, always capture provenance information. For this reason, how an individual special collection developed – closer to a public library, or closer to an archive – may affect its decision to collect provenance metadata. The data collected with this study supports this, as almost all institutions with overt ties to an archive provided provenance metadata.

In some cases, provenance metadata may be lost through the use of crosswalks (see "Provenance as Metadata"). When metadata is converted from another format, even

if provenance metadata was initially present, it may not have an equivalent in another standard, or the information may be spread over several fields. Provenance metadata may simply be lost in such a process. This problem is commonly cited for provenance when libraries convert their catalogue information into a union catalogue (Curwen & Jonsson, 2007). During this process, copy-specific information is considered of local or private interest, and is omitted from the union catalogue. This challenge teases at larger issues in metadata studies – the varied nature of metadata certainly presents flaws, and in its current forms, may not always be enough to adequately carry over information.

The presence of provenance metadata and the axes of provenance available within that metadata may also be affected by the time period information was collected. As described in "Provenance Theory", understandings of provenance have changed over time, and accordingly, information captured at a certain time may omit data. For example, for many years, evidence of ordinary people using books was not recorded in provenance metadata, which focused instead on famous people (Scheibe, 2010). Further, provenance theory has undergone waves in popularity. If metadata was captured when provenance was less popular, it was possibly omitted.

Even in cases where provenance metadata is offered, numerous metadata standards exist on which to draw upon as guidelines. This is evidenced by the large number of standards and elements used to capture provenance information discovered within this study. The reasons for this are partly historical, as explored in "Provenance Theory". Further, different standards seek to address different needs, and have therefore developed their own metadata schemes and thesauri over time, as explored in "Provenance as Metadata". Choosing a standard can be difficult, depending on the

emphasis a special collection wishes to focus upon. Provenance, like any metadata field, is therefore subject to great variation, even within the same library whose cataloguing processes may have changed over time. In the case of provenance, this is made even more complicated as understandings of provenance itself have also changed over time.

Similarly, even standards within the same community are always changing. This provides a notable challenge for special collections libraries (Zick, 2009), and often results in libraries with mixes of new and old standards. For example, in USMARC, the General Note field (500), established in 1973, was used for provenance information. A decade later in 1983, however, the Provenance Note field (561) was added. Many libraries continued to use the note field and did not switch over (Zick, 2009), likely due to issues of time and money. Alternately, an institution may introduce new records using the 561 field, without updating old fields, resulting in a mix of new and old standards.

Curwen & Jonsson (2007) sum this issue up succinctly:

> This is a continuing problem; we argue for improvements in the formats, but whether improved and new fields, however desirable, are adopted, and to what extent, remains in the hands of the individual chief librarians, their catalogues and their accountants. (p. 43).

Unless incentive is presented in revising metadata, this problem is likely to persist.

This study found that nearly 80% of digital special collections currently provide provenance metadata, a great increase over previous studies, and the literature agrees that "we do see today a growing understanding among librarians that provenance information is important," (Curwen & Jonsson, 2007, p. 45). This is reinforced by the creation of

"provenance" as a new element in many collections, reflecting the identified need for such a field in multiple collections.

Many scholars have observed, however, that "the framework for recording [provenance] has serious gaps" (Curwen & Jonsson, 2007, p. 45), and metadata standards have limitations making it difficult to catalogue provenance in a consistent manner (Lundy, 2008), as briefly touched upon in "Provenance as Metadata". For example, in standards which recommend capturing provenance in a "Note" field, often further guidelines are not supplied, such as how to structure information, or what context should be included. The variation of provenance axes discovered in this study reflects the results of these limitations, although libraries seem to be generally moving towards more complex descriptions of provenance, even if this is unstructured.

The large number of elements identified in this study support discussions in the literature which suggest that provenance information tends to be spread across several fields, or is relegated to local or note elements (Lundy, 2008). By mixing standards or creating their own field, many collections may circumnavigate issues inherent in cataloguing provenance, but create new problems for users in accessibility to data.

Another major challenge provenance encounters as metadata stems from the fact that in order to be useful for users, metadata needs to be visible and accessible. Provenance, however, is a complex theory, and as a result, tends to reflect rich information. Computer formats do not always display rich information well (Curwen & Jonsson, 2007). Likely for this reason, the use of general fields such as free-text notes or essays appeals to collections. Even if the information is unstructured (Curwen & Jonsson,

2007), the richness of provenance is more likely to be preserved. This study supports these suggestions, as most free-text metadata fields provided greater depth of metadata.

As a result of this formatting, however, finding provenance for a known book is easy, but looking for books with useful provenance is difficult.[20] This becomes markedly true when information within notes is not identified in any way.

Visibility of information can also present challenges in other ways. Some standards are embedded metadata, meaning that the information is not necessarily readily available to users, even if it is known. This parallels the fact that often, rare book and special collections keep separate files on provenance from the catalogue information (Lundy, 2008). This separation of information means it may not be visible or accessible in the online environment. The availability of such information may not even be indicated in the catalogue records. Users must instead contact each special collection individually to determine if such offline information is kept elsewhere. This prevents provenance metadata from being used cross-comparatively across institutions, or from being information that can be searched and linked to other relevant items.

Perhaps the major reason provenance information is often only available in analogue format and not within digital metadata is backlog. Backlog occurs when there are more items available to catalogue than time and workers to create the information for those items, or when special collections have catalogue information, but have not been able to digitize it due to financial restraints, lack of time, or technological barriers. Sometimes, collections will only digitize part of their catalogue to compromise with these

[20] Curwen & Jonsson (2006) note that in a search of the Library of Congress, using the advanced search for "Provenance Notes" and the term "Ex libris" yielded many missed results.

restraints, and often specialized metadata like provenance are left in analogue formats, as overarching institutional goals in metadata provision may aim towards broader audiences.

Special collections have long had problems with backlog, and this has become particularly true for archival materials since the nineteenth century (Buchanan, 2011). Several initiatives were launched in the 1990s to help special collections deal with institutional backlogs.[21] Although some collections may opt to use a code so that items of similar provenance can be quickly identified in the catalogue, such as the SPAC code (Buchanan, 2011), backlog still poses a significant challenge in catalogue provenance.

Backlog issues are complicated by the rise of big data. With increased numbers of digitized works, there is increased demand for means of analyzing data, creating a different kind of bottleneck. Lack of expertise with big data may add to this problem. Backlog may also affect linked data and interoperability, if an attempt is made to bridge with collections which are still experiencing delays in digital cataloguing. Backlog is also an ongoing issue for data curation and changing technology, which may require old catalogue records to be constantly updated to newer standards.

The cataloguing backlog experienced by many libraries makes "a rethinking of library operations […] timely, if not urgent" (Howarth, Moor, & Sze, 2010, p. 423). This reinforces the importance of this study, and the need to evaluate provenance metadata in digital special collections, and experiment with new ways of addressing these challenges.

---

[21] These included the National Historical Publications and Records Commission (NHPRC) Basic Projects programs, the Council on Library and Information Resources (CLIR)- Mellon Foundation Cataloguing Hidden Special Collections and Archives program, procedures carried out at research universities by William Jordan Patty, Carolyn White, and Tedine J. Roos, procedures carried out for rare books by Elain Bleckley Bradshaw, Stephen C. Wagner, Melissa A. Hubbard, and Ann K. D. Myers, procedures carried out at the National Archives by Ashley Bucciferro, general analysis conducted by Lynn C. Howarth, Leslie Moor, and Eliza Sze, and the creation of task forces and surveys such as the ARL Special Collections Task Force (2001 – 2006), and the OCLC Research Survey of Special Collections and Archives (2009 – 2010) (Buchanan, 2011).

Recommendations for Provenance Metadata Structure

Strong recommendations have thus been made by experts in the field for recording provenance, which would ideally be applied to cataloguing practices. Many of these address the challenges outlined above in capturing provenance as metadata.

Although provenance metadata is often incomplete and omitted for this reason, institutions should be encouraged to record incomplete information. This includes making note of handwriting, even if it is difficult to decipher (Curwen & Jonsson, 2007)[22], or leaving space within records for descriptions of provenance, such as a blank field (van Heel, 2011). Further, any provenance information present in embedded metadata, but not item records, should be made available in the item record.

By the same stroke, records should seek to provide more than just information on who owned an item. Provenance includes not only owner names, but dates of ownership, mode of transference, and locations the item was kept (Curwen & Jonsson, 2007; IFAR, n.d.). In fact, any information related to the path beginning with the origin of the book to its current owner, and the stops in between, is provenance (Curwen & Jonsson, 2007).

In 2007, Curwen & Jonsson noted that "[t]hese data are, however, seldom added to the catalogue records, and if they are, it is usually not done in a way as to make them systematically retrievable" (p. 32). In other words, it is not enough to simply add more provenance information if it has no structure and is recorded in broad fields. As noted above, this obscures searches and connections between items.

[22] The Folger Library actually suggested an entire field be created for handwriting in 2004, although none was ultimately created. Most guides suggest handwriting be recorded as a variant, currently (Curwen & Jonsson, 2007).

This study reveals, however, that a majority of institutions are actually looking towards including more contextual information in provenance metadata. This is a step forward in helping scholars. However, information remains recorded in an unstructured manner. Institutions need to structure provenance metadata more consistently.[23]

David Pearson (2005), one of the foremost authorities on provenance, has recommended levels of structure that institutions should choose to adhere to in general metadata fields, such as "notes". At the minimal level, Pearson recommends that owners and approximate dates be always included. At the higher level, however, additional contextual information should be included, such as locations, transcribed inscriptions (not summarized inscriptions), descriptions of bookplates, shelfmarks, and references to published sources on the collection. Pearson (1994) has actually recommended that new metadata fields be created specifically for inscriptions, bookplates, and armorial stamps, as these are key pieces of provenance evidence. More recently, the use of relator codes has been suggested, particularly in embedded metadata, to help highlight links between items, and to help create controlled terms for proper names (Curwen & Jonsson, 2006).

Further, Pearson has suggested that provenance metadata should allow users to "reconstruct the chain of provenance, as far as the evidence permits, and to make connections between different owners as appropriate" (Pearson, 1994, p. 320). This means that each step of acquisition and ownership should be clearly outlined in the metadata. Such a structure has large implications for cataloguing practices, and would require a means of connecting such data together in a dynamic way, allowing for

[23] No specific means of coding dates, prices, or places in connection with provenance are currently available in cataloguing standards (Curwen & Jonsson, 2007).

flexibility in the amount of data that can be added to a chain, in order for its productive use in digital tools and analysis. This may create new challenges in provenance metadata.

Although Pearson argues for a minimal level with only owners and dates, in recent years, there has been great emphasis on location within provenance metadata (see "Provenance Theory"), as part of increased sensitivities to the context of provenance. In use of MARC in particular, new codes were created to allow metadata to encompass specific places as listed in thesauri (Curwen & Jonsson, 2007)[24]. For example, cataloguers can specify Westminster, and not just London. In 2006, the Permanent UNIMARC Committee also accepted a proposal for the use of the new field 621 for Place and Date of Provenance (Curwen & Jonsson, 2007).

In light of this increased interest in location data within provenance, this thesis builds on Pearson's proposal for the minimal level of provenance metadata to include not only owners and dates, but also locations where data is available. These three axes are discussed further in "Provenance Theory", but inclusion of location data allows metadata to be visualized in new ways in digital mediums (see "Visualizing Provenance Metadata").

Another major recommendation is adherence to thesauri where possible. Many major databases exist, as well as thesauri, and no singular one is considered dominant. This leads to certain provenance databases being exclusive to certain collections. Thesauri are therefore not complete (Curwen & Jonsson, 2007). When collections use thesauri or databases, however, it is easier for computers to search records, and it also allows links to be made between records with the same terms (Scheibe, 2010). Databases

[24] Other MARC codes include: $m for geographic areas which are not states or cities, $e for buildings, $f for precise dates, $h for event names, and $5 for current item location (Curwen & Jonsson, 2007).

in particular allow different forms of provenance evidence to be documented in separate records, which can then simply be linked to the collection item (Scheibe, 2010).[25]

Budget and time are constant issues institutions face, and these hurdles are especially true for provenance which is based on copy-specific examination. Some cataloguers may not be trained in reading provenance evidence (Curwen & Jonsson, 2007), and it is not easy for institutions to simply backtrack and update their collections. In light of this issue, "a positive confluence between librarians and scholars, as they both seek to develop a methodology to identify the owners and provenances of books" (Bruni, 2011, p. 52) has occurred. In other words, there has been a shift towards more inclusive methods of creating provenance metadata. This generally entails researchers working in co-operation with institutions to create provenance metadata. Some have suggested the creation of web-templates or other easy-to-use tools for researchers, as well as continued collaborative work on guides and thesauri (Curwen & Jonsson, 2007).

Many of these suggestions can be interpreted as culminating in engaging with new tools to portray provenance metadata. This study has helped establish how provenance is currently being used in digital special collections. Some attempts have been made in collections studied to engage with metadata in new ways, particularly through the use of visualization.

In particular, both timelines and digital maps have allowed for copy-specific information to be displayed, alongside related browsing and searching. More than that, however, visualizations can help bring about many of the recommendations outlined

[25] Scheibe (2010) has described the use of a joint library network to create provenance authority records, called the Weimarer Model by the Arbeitsege Alte Druckemeinschaft at the Gemeinsamer Bibliotheksverbund as a potential model. This same network also allows collections to save time when creating records, and to make provenance facts available as soon as they are known.

above. In order to be useful within visualization, metadata needs to be structured and standardized. The more complete the metadata, the better the visualization. And axes of location and date are well represented through digital maps and timelines, arguing well for the inclusion of both those axes in the minimal level of provenance metadata.

Structured metadata and visualization can also be linked and searched, leading to "greater data interoperability within the cultural heritage community" (Elings & Waibel, 2007, abstract) for large amounts of data. Such metadata can also be read well by computers. Visualization further encourages researchers to participate in providing provenance theory to institutions, as results can be pictorially seen and more easily processed. The introduction of new tools also encourages institutions to update their metadata or to seek funding to reduce backlog in a way that brings prestige to collections, while bringing those same institutions into Next Generation Catalogue territory. The use of visualization is not widespread yet, although its presence in major projects like The European Library indicates we are on the cusp of productive use.

Now that the current state of provenance metadata in digital special collections has been established, the second half of the study can begin, which seeks to investigate the use of visualization for provenance as a means of addressing many of the inherent issues in provenance metadata.

CHAPTER SIX: VISUALIZING PROVENANCE METADATA

A two-part action research study was developed in order to address the research questions identified at the beginning of this thesis (see "Methodology"). The first half of this study, which examines the current use of provenance metadata, was presented in "Provenance Metadata in Digital Special Collections". The chapter presents the second half of the study, which seeks to examine the visualization of provenance metadata and its potential benefits.

Visualizing provenance builds on the findings of "Provenance Metadata in Digital Special Collections". Results suggested that provenance metadata is in a good position for visualization. The vast majority of collections offer provenance metadata (80%), and much of that metadata has at least three axes of provenance (owner, place, and date). Further, some collections have already experimented with metadata visualization and linked data. This suggests it is possible to use such tools productively in digital collections. Findings also reinforced numerous challenges cataloguers face when capturing provenance, including omission of metadata, lack of interoperability, unsure guidelines, multiple and changing standards, lack of visible metadata, and backlog. As suggested in "Provenance as Metadata", visualization may help solve many of these problems, while also meeting contemporary uses of metadata.

This chapter will begin by examining the relevant visualizations uncovered in "Provenance Metadata in Digital Special Collections" in the study on provenance metadata. This will then be supplemented with a brief environmental scan. From this and relevant literature, a set of criteria is developed to evaluate digital visualization tools. Five tools will then be tested and assessed using sample metadata gathered in *Appendix*

B, to ascertain the viability of provenance visualization. Data was collected and analyzed using the methodology outlined in "Methodology". Discussions and recommendations are also explored in this chapter.

Visualization in Digital Special Collections

Several different kinds of visualization were identified in the first half of this study. These included timelines, digital maps, and network visualizations. Each of these can be used to represent axes of provenance: time, place, and ownership respectively. Interestingly, no combination of visual metaphors was found in collections studied, suggesting that the full potential of provenance metadata has yet to be tapped into. An examination of these visual metaphors as used by digital special collections follows.

Examples of available timeline visualizations can be seen in both *Figure 10* and *Figure 11*. In *Figure 10*, the history of a specific item in the digital collection can be traced using the timeline visualization. This is useful for informing users about the item itself, although it does not tell much about related items, or further contextual information.

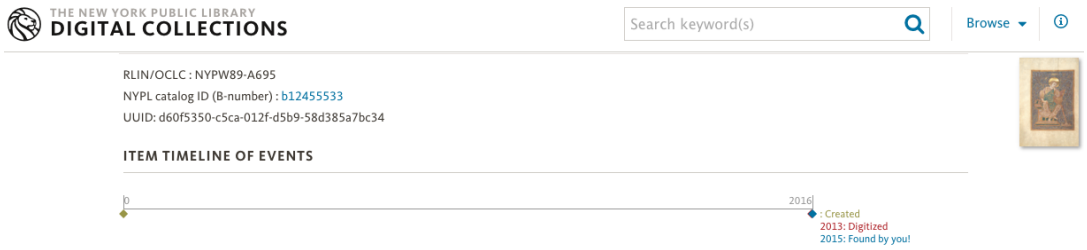

Figure 10: Timeline visualization from the New York Public Library Digital Collections (Manuscripts and Archives Division, The New York Public Library, 2015)

The timeline in *Figure 11*, however, allows users to browse for materials within the collection according to date on the timeline visualization. Users may not find out much information about a specific item, but they can easily browse for items by date. The

two visualizations, although using the same metaphor, are used in two different ways. One informs users with copy-specific information, while the other allows users to navigate collections. Timelines were the most frequently used digital metaphor in the digital special collections studied.

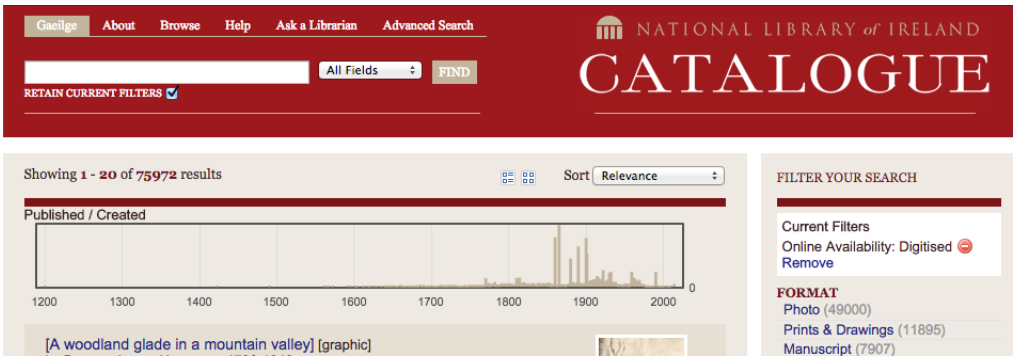

Figure 11: Timeline visualization from the National Library of Ireland Digital Catalogue (National Library of Ireland Catalogue, n.d.)

Another significant metaphor used by digital special collections was the digital map. Examples of the variety of digital map visualizations available can be seen in both *Figure 12* and *Figure 13*. These examples both draw from the same collection, the European Library; however, the metaphor is used differently in both cases.

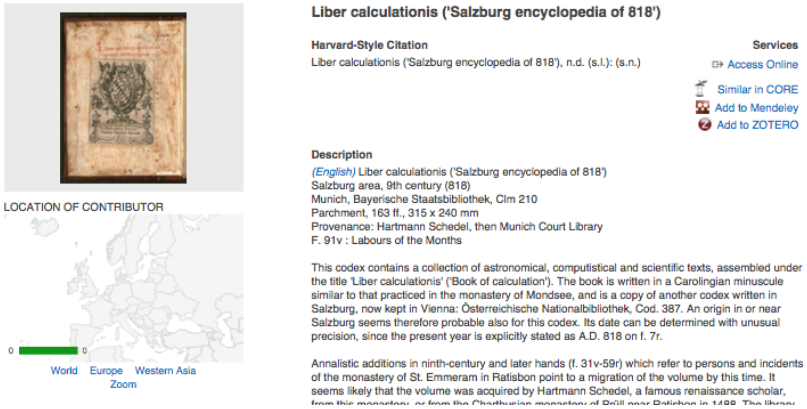

Figure 12: Digital map visualization from an item record in the European Library (Liber calculationis, n.d.).

In *Figure 12*, the digital map visualization can be identified in the lower left-hand corner. Here, the visualization offers users insight into the location of an owner or

previous owner of the item. The visualization is notably small, however, and does not lend itself well to browsing or exploring other items in the collection. Like the timeline in *Figure 10*, this digital map lends itself well to copy-specific information.

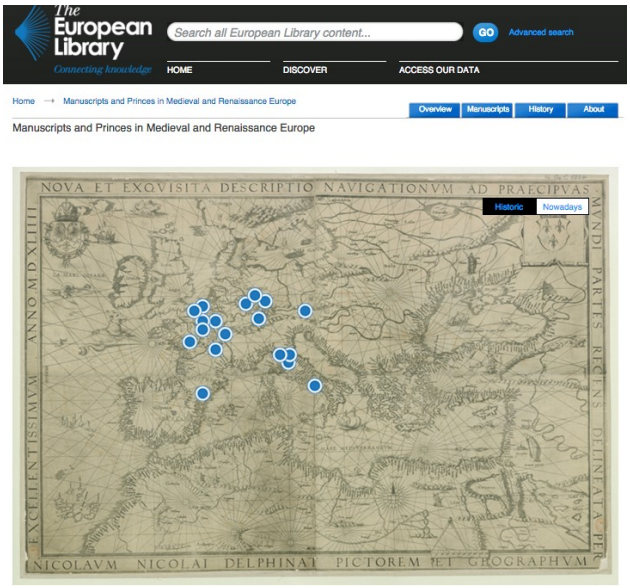

Figure 13: Digital map visualization for general browsing of the European Library Collection (The European Library, n.d.).

In contrast, however, *Figure 13* presents a digital map for navigating the digital collections according to location-based provenance information. Users can select items based on the location they are interested in, and multiple items are displayed at once.

Additionally, users can flip between the use of an antique map, or a more modern one, by adjusting the settings available on the top right-hand corner of the map. Although the same institution offers both visualizations, neither overlaps with the other. This means that users must backtrack to the visualization in *Figure 13* if they wish to browse, but to learn about an individual item, the context of other items must be lost, as in *Figure 12*.

Figure 14 illustrates the use of network visualization in the New York Public Library Digital Collections. Although the owners of the item are not illustrated in the network visualization, the broader collection the item belongs to, other items in that

collection, and the way the item is incorporated in the broader digital collection are illustrated, providing some network information for the item in question.

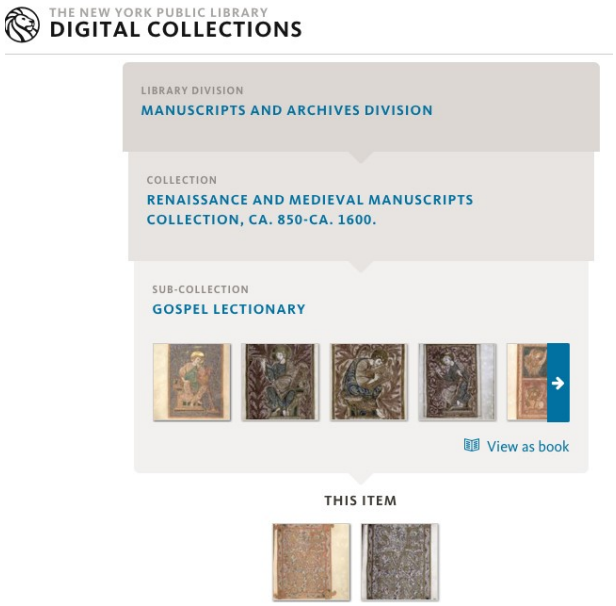

Figure 14: Network visualization from the New York Public Library Digital Collections (Manuscripts and Archives Division, The New York Public Library, 2015)

Network visualizations were not common in digital special collections, in spite of their history with library collections and metadata (see "Provenance as Metadata"). It is interesting to see how these three visualizations were all used independently, and often to portray some aspect of provenance: the history of an item, its location or origin, or the collection or potential historical ownership of an item. None of the visualizations, however, combine these intertwined facets together even though combination tools do exist. It seems that there is untapped potential in the use of visualizations to portray the multiple facets of provenance. Examining projects that combine visualizations is the next logical step in determining if provenance can be usefully portrayed to users this way.

Environmental Scan

Although it is helpful to understand the visualizations currently used in digital special collections, a supplemental environmental scan provides a fuller understanding of what current technology can achieve. An environmental scan identifies and analyzes projects with similar backgrounds and structures, in order to assess their strengths and weaknesses, as well as standards set forth in the industry.

There are literally hundreds upon hundreds of digital mapping projects. It was important to evaluate projects that had similar components relevant to the study at hand; in other words, evaluated projects had to have at least two of the three axes of provenance (space, time, and owners) represented. Eight projects fitting such criteria were reviewed: the Battle of Chancellorsville project, the Bookbinding Map, the Digital Literary Atlas of Ireland project, Google's My Timeline, Heurist: Roehampton in the 19th Century, Hypercities: UCLA, the Mapping the Republic of Letters project, and the Museum of Modern Art's Timeline and Map. The main findings for each project follow below.

The Battle of Chancellorsville: Scholarslab Project

The University of Virginia's Scholar's Lab has created many excellent digital mapping projects, although the Battle of Chancellorsville is perhaps one of the most interesting among them, due to the use of historical map overlay and the dual timeline available to users (see *Figure 15*). The project therefore combines two different metaphors productively: timelines and digital maps.

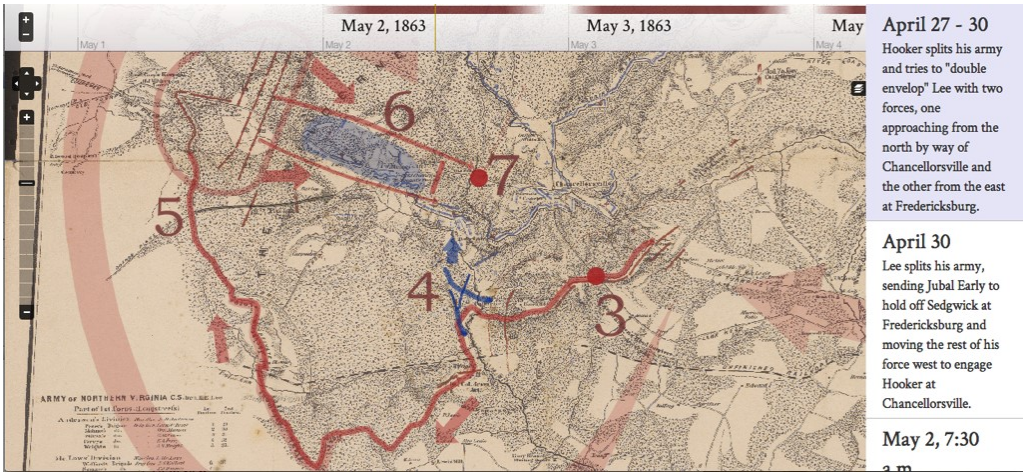

Figure 15: An image of the Battle of Chancelorsville Digital Map and Timeline Project (University of Virginia Library Scholar's Lab, n.d.)

Users can navigate different time periods using the overhead timeline bar, which will cause corresponding changes in the map below, or they can read information on the right-hand side bar about specific dates related to the project. Users are offered multiple modes of access in this design, which can provide new information. The map can zoom in and out, and events are ordered chronologically on the map. It notable that overlay of different timelines is not offered. Multiple "items" (or battles) cannot be viewed at once.

The Bookbinding Map

The History of Information's Bookbinding Map offers users the opportunity to read about major bookbinders by navigating a world map (see *Figure 16*). Clicking on a point on the map brings up information about the bookbinder who resided in that area.

This project does not feature a timeline display or demonstrate relationships or trade routes between bookbinders, but it does allow users to sort information by era and theme at the top of the interface. It is of significance to this study due to its content.

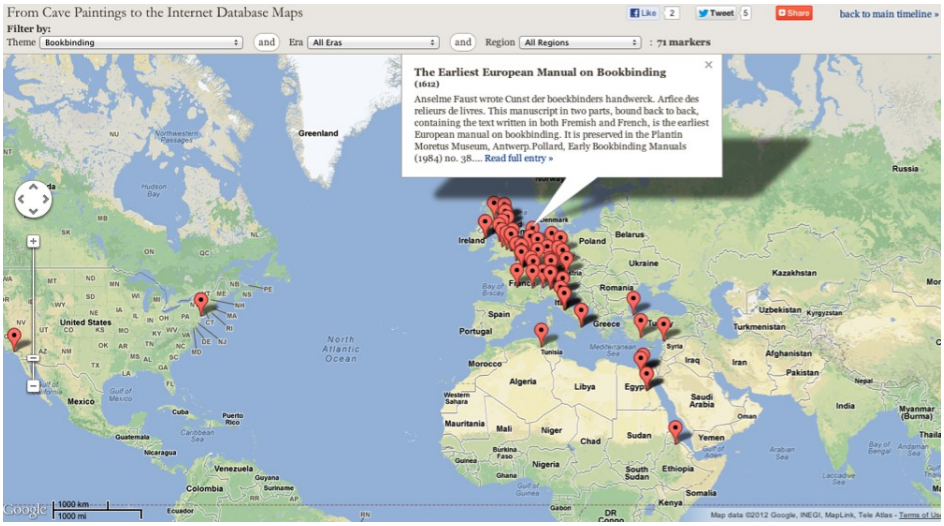

Figure 16: An image of bookbinding locations through the world, from the History of Information Project (History of Information, n.d.)

Digital Literary Atlas of Ireland

The Digital Literary Atlas of Ireland is a project that examines the works and lives of fourteen Irish writers between 1922 and 1949 (Trinity College Dublin, 2015), through interactive maps and timelines. It is generally aimed towards academics.

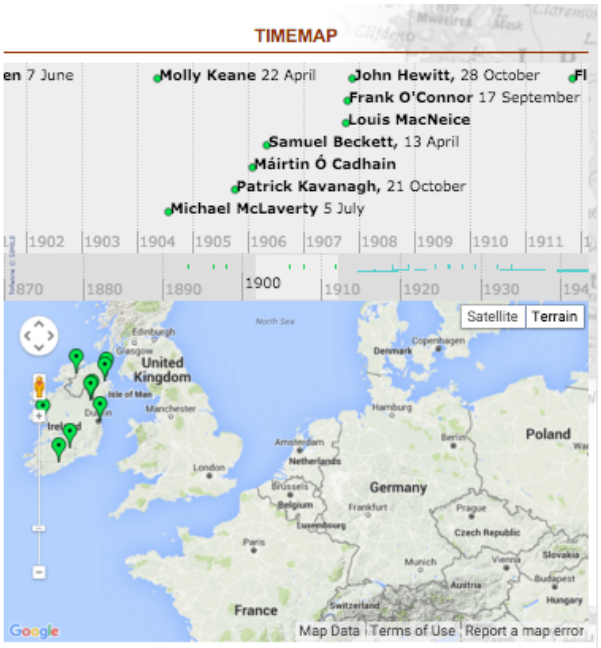

Figure 17: An image of the Timemap from the Digital Literary Atlas of Ireland (Trinity College Dublin, 2015).

The project can be navigated many ways. Users have the option of engaging with a Timemap, which features a timeline integrated with a digital map. When users move

along the timeline, corresponding points are highlighted on the map (see *Figure 17*). No connection is highlighted between these points, however. Clicking on an event or a location opens a small overlay with more information or related photos.

Users can also choose to focus on a particular writer or theme, opening individual maps or individual timelines. These more specific focuses, however, do not feature a timeline and a map that are linked together, and must be browsed separately.

Google's My Timeline

A note should be made of Google's My Timeline, a project that integrates timelines and Google maps. Although the focus of this tool is different from projects featured here, it is one of the most well-known geotemporal visualizations. My Timeline displays paths of a user on a digital map connected to a timeline that outlines events at specific locations. Data used to generate My Timeline is gathered from user cell phones. Photos taken on the displayed date and location are also available (admin, 2015).

Heurist: Roehampton in the 19th Century

The Roehampton in the 19th Century project was developed by Roehampton University, in London, England using the data management system Heurist. Heurist was developed at the University of Sydney, to help researchers in the humanities "design, create, manage, analyse and publish" (Heurist, n.d., para.1) databases.[26] With Heurist, users can create maps, timelines, network visualizations, tag their data, and support multimedia, among other uses. The Heurist website notes that it works with non-western systems, and dating uncertainty (Heurist, n.d.). The project uses a digital map with an integrated timeline to trace important events from the university's history (see *Figure 18*).

---

[26] Heurist was developed as a successor to the TimeMap tool, also created at the University of Sydney, and headed by the noted GIS researcher, Dr. Ian Johnson. (University of Sydney, 2016).

Many features of the Heurist tool would make it a good candidate for case study testing.[27] As the case study uses sample data, however, gathered from materials from a different institution, and the Heurist tool shares data on a large scale with other users, sample data was not deemed appropriate for testing the system. Instead, Heurist was evaluated through this example project.

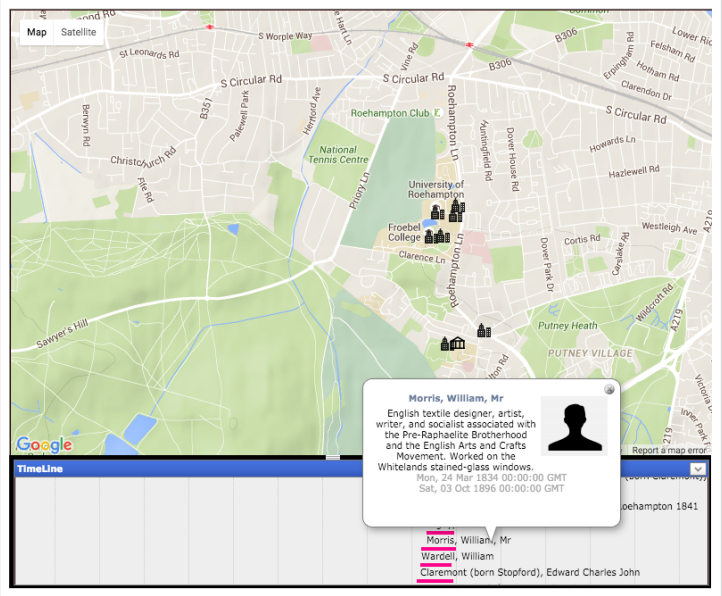

Figure 18: An image from the Roehampton in the 19th Century project (Roehampton, 2014)

As users scroll along the timeline at the bottom of the screen, different points on the map appear or disappear. The icon used for these points can be likened to buildings or human figures. Both the map and timeline can zoom in or out. If an event on the timeline is clicked, information is available in a pop-up bubble about the event. Users can also choose to search the database in a box below the map, which searches a more traditional layout of a timeline as a long list of information.

[27] Data uploaded to Heurist is stored in open source SQL databases, and can be used to create JSON, XML, and SQL archive packages without the need for programmers (Heurist, n.d.). Users must register an account to upload data to Heurist. This data is then made available to other researchers also affiliated with the management system. Heurist is well supported by documentation and FAQs.

Although the integration between timeline and digital map is well done in this visualization, the pop-up bubble supplied for reading about events is very small, and the connections between different points on the map is not highlighted in any way.

Hypercities: UCLA

One of the most celebrated timeline and digital map interfaces, Hypercities UCLA is a project based out of University of California, Los Angeles (UCLA). Users have the ability to choose between "collections" to display on the map, through a toolbar on the right-hand side of the map. This toolbar also features a "mini-map" of the world, allowing users to jump between related points to the displayed theme (see *Figure 19*). Users may also choose between satellite, street, and terrain views.

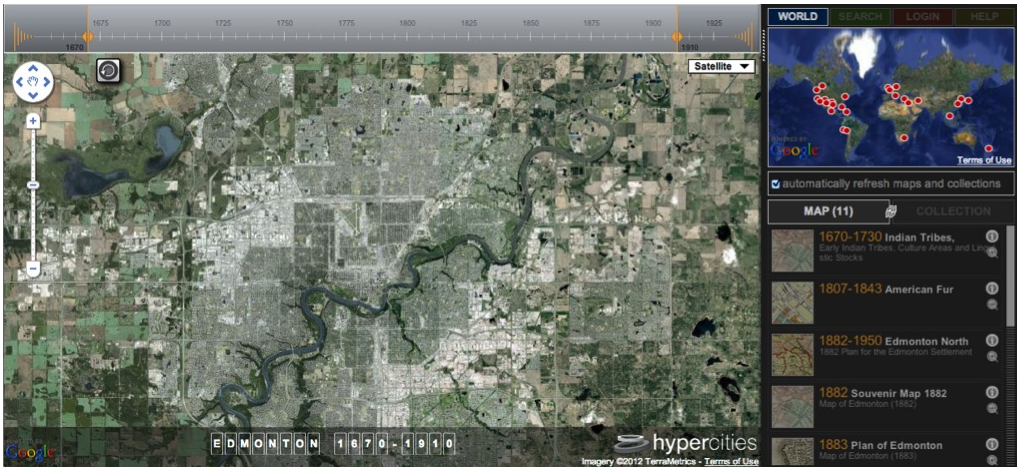

Figure 19: An image of Edmonton on the Hypercities Project (University of California, Los Angeles, n.d.)

Mapping the Republic of Letters

The Mapping the Republic of Letters project was born through a collaboration of international partners, including Stanford University, University of Oxford, Groupe D'Alembert, CNRS, and CKCC (Stanford University, 2013). The project presents numerous social networks of past scholars and scientific academics. These networks are created with "sophisticated, interactive visualization tools. [The project] also aims to

create a repository for metadata on early-modern scholarship, and guidelines for future data capture" (Stanford University, 2013, para. 2). Many of these goals are shared by the investigative questions of this thesis, making this project particularly interesting.

The tools used for this project were created by the Humanities + Design research lab from Stanford University. The most common visualizations include digital maps (see *Figure 20*), focused on a specific time period. Data is then visualized according to different categories of publication data, selected by the user. Many different categories may be used at once on the same map. For example, users may select "Actual Places" within the metadata to be displayed, as well as "False stated places", which refers to places of publication that were incorrectly stated in an edition.

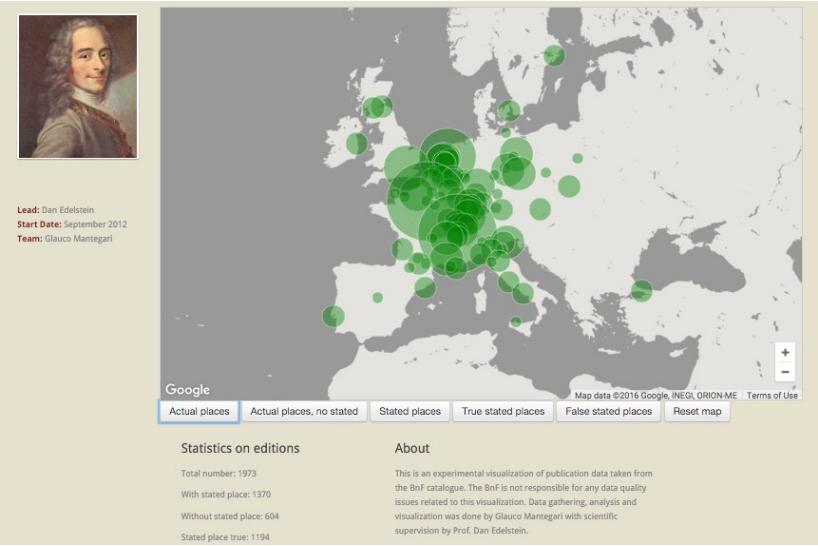

Figure 20: An image of Voltaire's correspondence (1712-1800) on the Mapping the Republic of Letters Project (Mantegari & Edelstein, 2012)

Locations on the map are identified through the use of different sized circles, as opposed to specific points, corresponding to the number of publications in a place. Clicking on a circle allows users to see the name of the location, and the number of publications associated with it.

The second visualization used by the Mapping the Republic of Letters project combines digital maps, histograms representing time, and alluvial or Sankey diagram views representing relationships, emphasizing connections between data (see *Figure 21*). Users can click on a point on the map, a bar on the histogram, or a section of the alluvial diagram, and metadata about related items, such as the author, recipient, date, and URLs linking to the catalogue record will be made available.

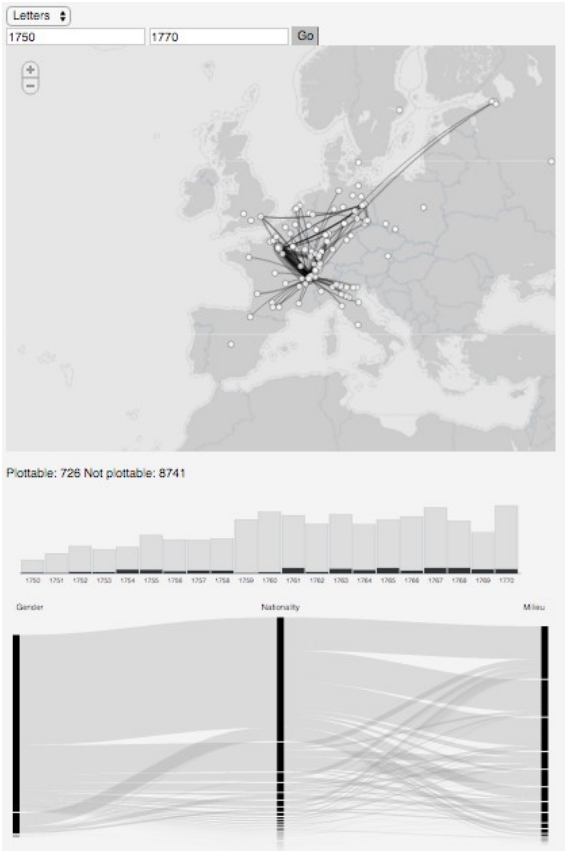

Figure 21: An image of Voltaire's correspondence visualized as a map, a histogram, and a Sankey diagram in *Ink* from the Mapping the Republic of Letters project (Design Humanities, n.d.)

It is of note that although the diagrams are linked in this second visualization, the map is representative of a segment of time. To view a different date, the parameters of the visualization must be adjusted. Multimedia such as images is not supported.

Museum of Modern Art (MoMA)'s Timeline and Map

The Museum of Modern Art (MoMA) in New York, New York features a timeline and digital map visualization, which traces the history of the institution's traveling exhibitions. Users can browse either by clicking on images along the timeline at the bottom of the visualization (see *Figure 22*), or by clicking on red dots on the map.

Once an event has been selected, related points on the map are highlighted in blue. Information about the event appears under the timeline and the map. Other images or related multimedia to the event are also available. Users have the option of selecting the type of information they would like between exhibitions, events, programs, and news.

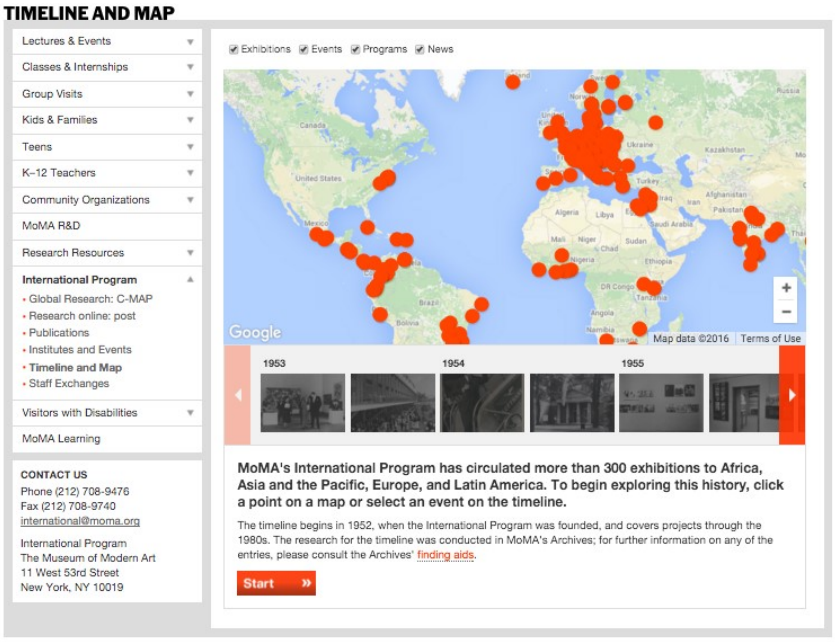

Figure 22: An image of MoMA's Timeline and Map (Museum of Modern Art, 2015)

Visualization Criteria

In "Provenance as Metadata", provenance was shown to lend itself well to the five fundamental categories of Ranganathan's faceted classification scheme, PMEST. These facets include where a book has been, when it was in a place, copy-specific features of a book, how a book changed hands, and the physical material of the book itself. Because

provenance is a multi-faceted concept, it should lend itself well to visualization. Inclusion of as many of these facets as possible should be a baseline criterion for any visualization.

Additionally, Chickering & Yang (2014) have compiled a list of criteria that any discovery tool used in Next Generation Catalogues (NGCs) should have. These include federated searching, flexible interfaces which appeal to users, enriched content (e.g. images, etc.), faceted navigation (by location, date, authors, etc.), spell-check, auto-completion, recommendations, and relevancy rankings, web 2.0 features such as social media and potential for user contribution, and persistent links.

This list of criteria can also be informed by the environmental scan, which shows which user-beneficial features are possible to create. Drawing from the environmental scan, ideal provenance visualizations would allow users to choose between the display of different collections or materials or eras; choose different map views (e.g. antiquated maps); see the chronology between points on the map for a specific item; and quickly understand connections or relationships between collections, items, events, and locations.

Using these criteria as a starting point for evaluation, mock data was entered into identified software to determine the best tool available to display provenance metadata in digital special collections.

Visualization Tools

Five digital tools were identified for testing with sample data (see *Appendix B*). All the tools focus on the combination of time and place, with the ability to incorporate narrative and relationships. Surprisingly few tools combining visual metaphors are currently available on open-source platforms, but those identified for study have strong ties to the information retrieval community, making them relevant to this thesis.

It is of note that several prototypes and digital tools for visualizing geospatial metadata exist,[28] although due to the precise nature of the scientific metadata used in those communities, and associated costs, these tools were not considered for this study.

The five digital tools tested in this second half of the action research study were

1. Neatline, an Omeka plugin, which allows research and special collections libraries to create interactive timelines and maps (Scholar's Lab, n.d.a),
2. Palladio, a web-based platform for visualizing complex, multi-dimensional data using maps and network visualizations (Humanities+Design, n.d.b),
3. Storymap JS, a JavaScript tool that tells stories as a series of events on a map, integrated with multimedia (Northwestern University Knight Lab, 2015),
4. TimeMap, a tool developed to combine the popular digital humanities interactive SIMILE Timeline with Google maps (Google code, n.d.), and
5. TimeMapper, an open-source project created to help users easily create timelines and maps online (Open Knowledge Foundation Labs, n.d.).

Findings

Over the course of the short case studies, each one of the tools was assessed on the criteria outlined above, and the tools' abilities to represent the axes of provenance. A number of tool limitations, as well as recommendations are made after the case study, including a suggestion for the best tool currently available to digital special collections.

Neatline Project

Neatline is a plugin designed for Omeka, an open-source content management system largely used by libraries and cultural heritage institutions. Omeka was developed

[28] Examples include GeoPWProv, which uses maps and faceted metadata to generate visualizations (Sun, Yue, Hu, Gon, Zhang, & Lu, 2013), MetaViz, InfoCrystal, VisDB, and TileBars (Göbel & Jasnoch, 2001).

by the Roy Rosenzweig Center for History and New Media at George Mason University (Omeka, 2015), and focuses on online digital collections. It allows institutions to exhibit cultural heritage items on the web in dynamic ways, while adhering to the Dublin Core metadata standard (Omeka, 2015). Omeka has won a Mellon Award for Technology Collaboration from the Andrew Mellon Foundation (Omeka, 2015).

Neatline was developed by the University of Virginia Library Scholar's Lab, and is purposefully geared towards "use in research and special collections libraries and with scholarly digital projects that build on library or museum-managed archival resources." (Scholar's Lab, n.d.a, para. 1). It seeks to offer users the ability to tell stories with interactive maps, timelines, and narratives. Many excellent digital mapping projects have been created using Neatline, such as the Battle of Chancellorsville (see above). Literature suggests Neatline helps visualize the unknown, and encourages researchers to disrupt, reorder, and practice new forms of inquiry (Earhart, 2014).

The Neatline plugin, which primarily creates digital maps, can also be combined with numerous other plugins that offer different features. Of interest to this project is NeatlineSimile, which combines the SIMILE Timeline[29] widget with Neatline exhibits (Scholar's Lab, n.d.b), so that events may be placed in time as well as place through the generation of digital timelines alongside the map. Users can also choose to combine Neatline with the Wayfinder plugin. Wayfinder allows creators to order events chronologically, so they must be viewed in a certain order. Both the SIMILE and Wayfinder plugins were tested in the case study.

---

[29] SIMILE Timeline itself is a tool that was created as a joint cloud-based project between the MIT Libraries, and the MIT Computer Science and Artificial Intelligence Laboratory (Marquez, 2012). The SIMILE Timeline is a celebrated digital humanities tool, and several example timeline projects exist.

In order to use Neatline, Omeka must be installed. This makes Neatline an ideal tool for those collections already using the Omeka platform. Neatline is best installed by users with some degree of knowledge in coding. As Omeka serves as a basis for Neatline, metadata is structured according to the Dublin Core standard.

Once Neatline has been installed, users have the ability to create exhibits with the plugin - essentially, digital maps. An indefinite number of exhibits can be created. When an exhibit is created, users can enable extra plugins, such as NeatlineSimile and Wayfinder. When creating the visualization, users can select a default map from preset options, and maps for alternate views, such as street maps, satellite maps, terrain maps, etc. Users could alternately enable a base map from another website, or use GeoServer in co-ordination with Neatline to create more complex maps.

Once an exhibit has been created with plugins and base maps selected, users can begin to add records. Exhibits may be populated with individual items or entire collections from Omeka. Similarly, items can be added one at a time, or in bulk. Neatline provides excellent documentation to help users set up exhibits and create records within them (Scholar's Lab, n.d.a). It is of note, however, that although multiple items may be added to an exhibit, if each item has several events associated with it, a new record must be created and added for each event. This can be a time-consuming process.

Users may add descriptions for events, as well as a title, tags, and a link to its Omeka record, which automatically displays Dublin Core metadata associated with the item. Multimedia such as images, audio, or video can also be included. The event can be added to the timeline, the digital map, or as a Wayfinder point, or any combination

thereof. When users click on an event, associated metadata will be displayed, as well as the event's location on the map and place on the timeline (see *Figure 23*).

Timespans can be included on the timeline by entering start and end dates, or a single date may be entered for an isolated incident. The timeline is also customizable in its display and size. Users may choose to display dates over years, decades, centuries, etc.

In order to display an event on the map, users must search for the location, and then can create a polygon or a point over the appropriate location. As any colour, shape, or size can be used to pinpoint a location, there is a great deal of flexibility in geographical display. If a location is used more than once in an exhibit, the level of opacity of the area increases, to help illustrate this. Very opaque areas on the map are signs of well trafficked areas. Users can choose what order layers are displayed, particularly with the Wayfinder tool, and can also add lines to the map. Lines between locations can be helpful in displaying connections between locations or item journeys.

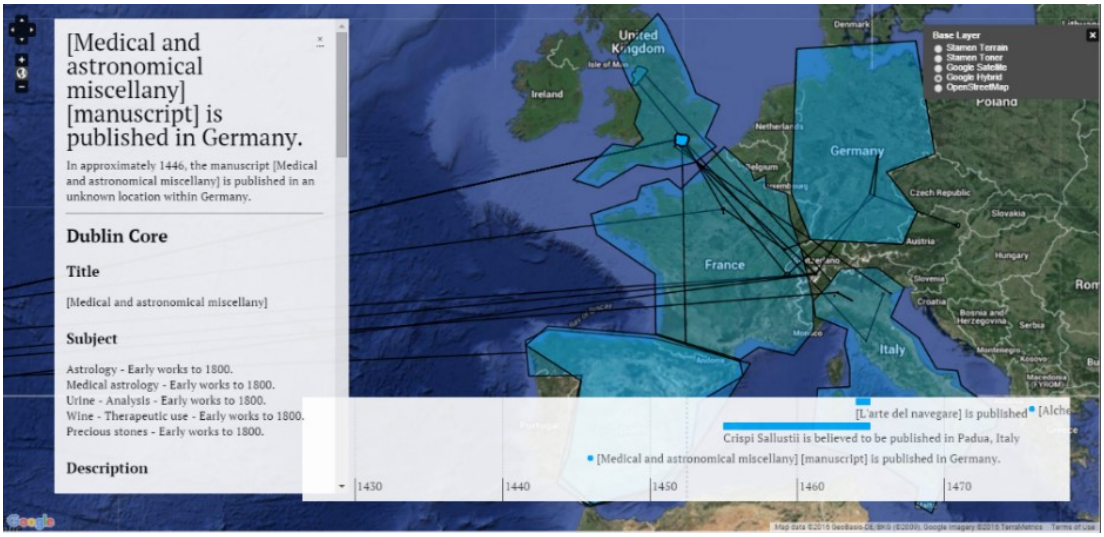

Figure 23: An image of sample data visualized in Neatline with only the Simile plugin activated

Enabling the Wayfinder plugin allows users to order events in the chronology they would like end-users to see events (see *Figure 15*). Although this is helpful in

exploring the journey of a particular item, it also prevents the inclusion of multiple items on the same map or collection, as there no way to differentiate between item histories. For this reason, the Wayfinder plugin is not recommended for visualizing provenance.

Neatline is a powerful tool for visualization. It is extremely customizable on several levels, and demonstrates ambiguity inherent in provenance well. For example, if a date is unknown, users may omit including a date, or may use a best estimate and explain further nuances in the description. Alternately, a place may be omitted if it is not known, or a larger geographic area that is known, such as a country, may be captured.

There are also limitations to the tool, however. Adding an item several times to create new records for each event associated with an item is a frustrating and time consuming endeavour, particularly if the user is working with many materials and events. If the Wayfinder plugin is enabled, then more than one item cannot be accommodated at once. Although several items can be displayed at once in cases where Wayfinder is not enabled, it is not always easy to differentiate between them, even with the use of lines. If many items are represented, the lines may become confusing, especially as there is no way to highlight which lines belong to which items.

When users input information in the Neatline tool, it can be frustrating searching the map for a particular location, as there is no search-bar for such information. It is also impossible to indicate a time span that continues "to present" on the timeline - instead, users must guess which date to include in the future. If an exhibit contains a large amount of data, it is also slow to upload or edit further information in it.

Perhaps the largest limitation in using Neatline lies with end-users however. In spite of the availability of tags, there seems to be no way of differentiating between

collections on the map. This is made even more problematic by the fact that users have no control over what they see on the map - they cannot search the records or select what to display. Instead, end-users can only use an exhibit as decided by the collection.

Palladio

Palladio is one of a number of open access visualization tools developed by the Humanities + Design research lab at Stanford University, which also include Breve, Grand Tour Timechard, Idiographics, Ink, and more (Humanities+Design, n.d.a). Both demos and code are publicly available for these tools. The tool Ink, created in 2012, was used in the Mapping Letters project, discussed in the environmental scan above.

Initially created in 2014, Palladio is a tool in ongoing development that allows users to examine complex and multi-dimensional data through digital maps, graphs, and timeline visualizations (Humanities+Design, n.d.a). Although Ink and Palladio are similar in nature, Palladio was chosen for testing as it better fits study criteria. To some degree, as a more recent and complex tool, Palladio is a successor to the Ink visualization.

The Palladio tool was created through the support of the Office of Digital Humanities within the National Endowment for the Humanities (NEH), the Vice Provost for Online Education at Stanford, the Wallenberg Foundation, the Stanford University Libraries, and the Dean of Research at Stanford (Humanities+Design, n.d.b). Anyone can use the tool – it is not code that needs to be downloaded and used by someone with experience, and no account is necessary to engage with the tool.

To use Palladio, users must create a CSV sheet that can be uploaded or copy-and-pasted into the tool interface. This data must then be classified by data-type (text, date, co-ordinates). The use of CSV sheets can be a limitation, unless data is already in this

format. If a collection has many items with complex metadata, converting information into a table is very time consuming. Further, complex materials will need many columns. Creators must also carefully design tables so information is appropriately represented in visualizations. This is not an intuitive process, and Palladio offers little support documentation.

Creators must also invest time into determining GPS coordinates in latitude and longitudes for every location they wish to include on the digital map. Palladio offers no means of helping creators discover what those coordinates might be.[30] As points are generally used on the maps, geographic coordinates may be misleading, or too specific to accurately represent the provenance metadata. Ambiguity is difficult to portray in the data. Special characters are not accepted; so unsure dates cannot be represented through traditional dashes or question marks. To that end, timespans cannot be represented in the data either. This may lead to the omission of dates, or the creation of false data.

Other issues creators may encounter in uploading data include the fact that data cannot be edited in the tool itself. Instead, the creator must correct the original CSV sheet, and then re-upload it and re-categorize columns according to data type. The tool does nicely highlight errors within the metadata, however, by colour-coding issues in red.

Once the uploaded data has been cleaned and categorized, it can be used to generate visualizations. Creators can choose which type of visualization to customize first, through navigating the tabs at the top of the page, and adding layers or customizing the visualization (see *Figure 24*). These tabs include: Map, Graph, Table, and Gallery.

[30] In this case study, coordinates were retrieved using http://www.gps-coordinates.net/

In the Table and Gallery views, creators can choose to display data as a list or a grid respectively. In the Gallery view, items can be linked to multimedia such as photos, or to outside links. Summary information for the item, as well as other metadata such as authors, dates, etc. can also be represented in the grids. In the Table view, data is portrayed in a list or a table, but creators can choose which columns to include in the table, and the order in which those columns appear.

In the Graph visualization, creators can visualize relationships between two dimensions of data, as nodes connected by lines. This visualization is limited by the fact that only two dimensions of data can be linked together, meaning that some axis of data will be left out. For example, Titles can be linked to Provenance Locations, but cannot be linked to Provenance Dates or Owners as well. Further, if provenance metadata has been broken up into several columns (e.g. Owner 1, Owner 2, Owner 3) so it is easier to manage, then only one of those columns can be chosen. This limits the functionality of this visualization, although it is aesthetically beautiful.

Figure 24: An image of sample data visualized in the Map facet of the Palladio Tool (Humanities+Design, n.d.b)

The Map visualization, most pertinent to this case study, allows GPS coordinates from the data to be represented as points on a digital map. Lines can be used to show relationships between points (see *Figure 24*). Creators can switch between different types of terrain and point colours, but only while designing the tool – end-users cannot individualize maps in this way. Maps zoom in and out for both creators and end-users, however. The Palladio tool mentions that points can be sized to be relevant to the data (Humanities+Design, n.d.b), although this requires creating geoJSON files customized to the area – another time consuming process that requires some technical specialization.

When end-users hover over a location, related lines are highlighted, helping point out related places (as long as the provenance locations are all contained in one column, or a new layer must be created for each location and only the immediately preceding location is highlighted). There are many limitations however – if a point is repeated many times for different objects, only one object will be portrayed. As well, it is difficult to manipulate the table data in the many layers required to build the map, a fact that would be only more difficult with more than the ten objects used in the case study. Most importantly, however, users cannot see information about the item outside of the title, so the points provide little context to users – what city is represented? Where is this location in the overall journey of the book? What item is represented at all?

Palladio does offer users some timeline visualizations (see *Figure 24*), although this is notably only as a filter, not as an interactive timeline that can be browsed and tied to other visualizations through browsing.

Once creators are satisfied with their visualizations, they can export the information as SVG files or JSON files for use in their own projects. The tool is notably

frustrating to use, as creators cannot save as they work. If the "Back" button is accidentally pressed, work is lost, and being unable to save can also make it harder to manage large amounts of data. Although some documentation is offered to help users navigate the tool, this site seems to go down frequently, making it unreliable. Palladio offers an interesting means of highlighting and visualizing relationships within data, however, even if the tool does not lend itself well to the complexities of provenance.

Storymap JS

The open source tool Storymap JS comes from the Knight Lab at Northwestern University, and operates under an MIT license[31] (Northwestern University Knight Lab, 2015). It was designed to allow users to create a series of events tied to specific locations, that when presented together portray a narrative. The tool can be used in conjunction with large photographs and other multimedia (Northwestern University Knight Lab, 2015). The development of Storymap JS officially began in 2013, and ran until 2015. The tool was designed using JSON. An extensive FAQ section is available on the Storymap JS website (Northwestern University Knight Lab, 2015), and further help is available to users through contact information to a support team for the tool.

Using the Stormap JS tool is very easy. Projects are initiated after login with a Google account – no additional passwords or other information are necessary. After logging in, users can create either map projects or image projects. Projects are easily saved to the user's account in this system, and can be returned to multiple times without

---

[31] An MIT license is a permissive and free software license originating from the Massachusetts Institute of Technology. This allows anyone with the software to deal "without restriction, including without limitation the right to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software" (as cited in Rouse, 2011) as long as the copyright and permission notice is included in the software (Rouse, 2011).

information being lost. After assigning a title, users can provide information about the project, and create “slides” in a specific order to capture events. Once completed, a project can be embedded in another site through provided code, or a link can be used to direct end-users to the project. Creators may also share the project via social media.

Creating slides is simple. After creating a slide title, users can add a URL, multimedia such as images, and further descriptive information. To place the slide on the map, users simply have to search for the location, and a point on the map will be added. A proper timeline is not displayed alongside the digital map, but the ordered outline of the slides gives a sense of the order of events. Further information on dates and owners can simply be added to the description box. Colour schemes, map terrains, and zoom level can also be modified at this stage. Although creating slides is a simple process, large amounts of data cannot be uploaded at once to the tool.

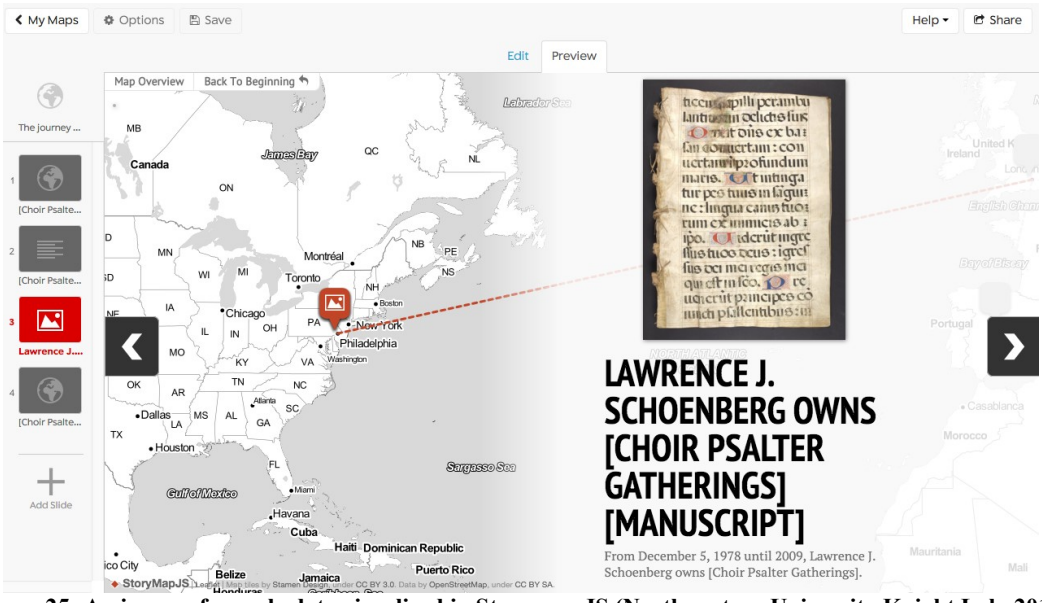

Figure 25: An image of sample data visualized in Storymap JS (Northwestern University Knight Lab, 2015)

The end product of a Storymap JS project is aesthetically pleasing and easy to understand (see *Figure 25*). End-users can click on any point, and instantly understand where that point falls in the overall journey due to the highlighted slide on the left-hand

side of the screen and connected doted lines on the map. Multimedia and descriptive information is instantly displayed. It is also easy to see what event and location immediately precede the point as the path becomes highlighted in red. When examining the overall journey, a small mini-map is also available to end-users.

Noteworthy drawbacks of Storymap JS include the display: information is gathered all in one small paragraph, and cannot be read in a side window. Although creators can change maps and terrain, end-users do not have this option, and using the map zoom is not intuitive. Although ambiguity in ownership and time can be easily addressed in the description, place ambiguity is harder to demonstrate, as every event must be tied to a location.

The largest limitation of the Storymap JS tool, however, is the fact that each item has to have its own slide project – multiple items or journeys cannot be displayed on a single map. This limitation prevents Storymap JS from being a good option for prospective rare book libraries seeking to visualize provenance metadata. Storymap JS is a notable tool; however, due to its intuitive means of creating events in an items journey, the ease of saving and sharing projects, and its clarity in demonstrating an item's journey.

TimeMap

The open-source TimeMap is a javascript library that combines the SIMILE Project with online maps, in order to create a functional geotemporal system. Created by Nick Rabinowitz in 2011, TimeMap is now somewhat dated, although it remains an extremely influential humanities tool due to its versatility and relative ease of use. Notably, projects such as Heurist (Heurist, n.d.) and the Canadian Writer's Research Collaboration (CWRC) have developed or are in the process of developing similar

SIMILE Project timelines with Google Maps (GitHub, 2015). At one time, TimeMap was also available as an application that could be integrated into Drupal, although it has not been supported for several years now (Drupal, 2014). As these tools are not available for case studies, and considering the importance and influence of the TimeMap tool, TimeMap was tested with the sample data in this study in spite of its dated nature.

In order to make use of TimeMap, some programming experience is required. Users can download the TimeMap code from Google Code. The code is open source, under an MIT license, and has been downloaded over 5,000 times (Rabinowitz, 2016). Data can be loaded onto TimeMap as inline JSON, remote JSON, or KML. It is simultaneously displayed on a map and timeline. As users scroll along the timeline, information will appear or disappear on the map, so that points portrayed match the timeline. If users click on a point, then a pop-up will appear, where information, multimedia, or links can be displayed to end-users (see *Figure 26)*.

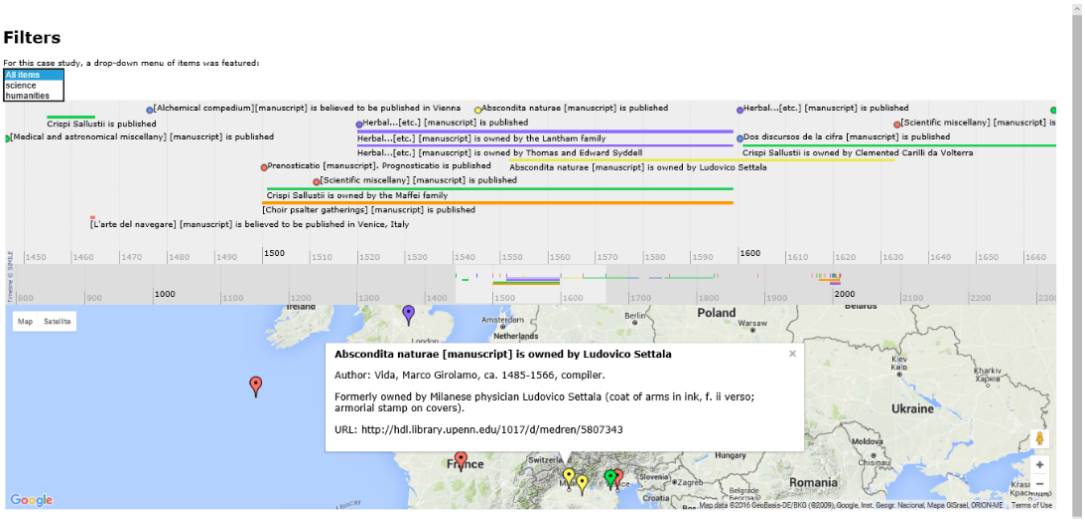

Figure 26: An image of sample data visualized using TimeMap

Developers are granted a large amount of customization in how information is displayed depending on their needs and coding skills. For example, the dates shown on

the timeline can be adjusted, and users can choose to display intervals on the timeline according to the timeframe they prefer (days, years, decades, etc.). Points on the map, which are assigned using latitude and longitude, can also be displayed as polygons, allowing for flexibility in unknown locations. Both locations and dates can be displayed on the timeline without a corresponding date or location, allowing for expression of uncertainty, especially when enhanced with information from description boxes. Relationships can be demonstrated between points using lines, which appear or disappear as points do when moving along the timeline, or by adjusting the code so points are shown progressively. Markers can be customized for colour, creating potential for colour-coding. For example, uncertain dates could be colour-coded differently than other dates.

Other helpful features can also be developed for users. For example, toggle boxes can be introduced for individual items or collections, which is useful for displaying large amounts of data as it can be grouped and displayed many different ways. Filters can also be added, so end-users can choose from a list what they would like displayed. Maps can also be somewhat customized – not only in terms of zoom or terrain, but whether or not the map is drawn from Google Maps or other providers, such as Open Layers (Rabinowitz, 2016). In fact, the size of the timeline, and the size of the map themselves are easily customized. Some of these features have been illustrated in *Figure 26.*

The fluidity and flexibility of the TimeMap software are impressive. They are notable advantages of using code directly as a tool. TimeMap faces limitations for selection as the best tool in spite of this versatility, because it is dated in face of changing coding standards. As happens with some open source projects, it is likely not maintained anymore. TimeMap also requires the greatest familiarity with coding of any tool

examined in the case studies, making it difficult to use for smaller special collections lacking access to personnel with the necessary skills to employ the application. It is also very time consuming to modify and input data.

Timemapper

Timemapper is a digital tool developed by the Open Knowledge Foundation Labs (Open Knowledge Foundation Labs, n.d.). Freely available to anyone on the Internet, Timemapper uses TimelineJS, ReclineJS, Leaflet, Backbone, and Bootstrap. The code can be downloaded online (Open Knowledge Foundation Labs, n.d.). In order to use Timemapper, however, no coding skills are required.

The first step to using Timemapper is to create a Google Spreadsheet containing all the information you would like to include in the TimeMap. Timemapper provides a template to help users get started. Every event has its own row in the spreadsheet. Media and URLs can also be added. Locations on the map are drawn from latitude and longitude, or GeoJSON, which must be added to the spreadsheet. This is very easy to do however, as the template provides a formula that automatically looks up and fills in GPS coordinates by simply writing out a place name. Once the Spreadsheet is prepared, users must publish the spreadsheet to the web, and copy the Spreadsheet URL. Pasting the URL into the Timemapper browser, filling in form information (such as selecting how dates are portrayed in the spreadsheet) and pressing "Publish" generates a TimeMap.

The visualization shows a map on the right-hand side of the screen, which can be zoomed in or out (see *Figure 27*). Users do not have the option to change terrain on the map. Points are illustrated with blue markers, and clicking on a point on either the map or the timeline brings up the item in a large box on the left side of the screen. Here, links,

media, and descriptions of the item can be found. The timeline is found on the bottom of the screen, and can be navigated by scrolling both left and right. The visualization is perhaps one of the most aesthetic evaluated in the case study.

Figure 27: An image of sample data visualized using the Timemapper tool (Open Knowledge Foundation Labs, n.d.)

Timemapper allows for some flexibility – although colours cannot be formatted, and the timeline dimensions are also set, locations can be included without dates, and vice versa, allowing ambiguity in location and time to be portrayed. Older dates are in slightly faded colours compared to newer ones on the timeline. A notable limitation is the inability to show relationships between locations or events in the visualization, however.

Although users are not required to, an account can be created by linking to an existing Twitter account, in order to save the TimeMap, and return to it for editing. Once generated, the map can be embedded on another site, or the page itself can be shared. Users can also come back and edit their original Google Spreadsheet, and changes will automatically reflect in the visualization. Large amounts of information may take a long time to load, however. This tool is the easiest to use of those evaluated in the case studies.

Tool Limitations

All of the tools studied in these short case studies exhibited limitations, as discussed above. The largest limitation common to the visualizations is the ability to display large amounts of data – when hundreds of items are added to the tools, it will be difficult to navigate the timeline and the map. Only one tool, the TimeMap, offered a possible solution to this problem, through the use of filters and the ability to select which collections and materials to show. This ability to filter by collection or owner makes TimeMap the only tool to successfully integrate some form of network visualization.

The facets of space, time, and personality were well represented in all tools. Digital maps, timelines, and the ability to provide descriptive information about owners were offered in all the tools to some degree. The two weakest tools in the studies were Storymap JS, where multiple objects could not be included, and Palladio, where ownership information could not be simultaneously shown with time and place. Most tools also allowed for the facet of matter to be represented through support of multimedia and enriched content, such as images, relevant links, and additional text. The facet energy was available through the use of descriptions, where transactions could be described.

Drawing on the criteria suggested by Chickering & Yang (2014), no tested tool offered federated searching in their interface, although the Heurist tool from the environmental scan offers limited searching. By extension, no tools offered spell-check, although TimeMap allows the use of drop-down menus and filters.

Most tools offered some means of faceted navigation, usually through time or place. StoryMap JS even offered mini-maps for quick navigation. In most cases, therefore, there is some flexibility in interfaces. Tools were mixed in their ability to use

social media – Storymapper and StoryMap JS allowed direct use of social media, while most other tools only offered persistent links, which could then theoretical be shared.

Some significant limitations in the tools can also be identified through the criteria developed in the environmental scan. Most tools allow for the customization of maps as desired by the user, with the exception of Palladio, TimeMapper, and Storymap JS, although the last two do allow some flexibility between terrain and street maps. Zoom was an option in all tools. The largest limitations however, were found in those tools that could not demonstrate multiple items, such as Storymap JS, or that had an inability to highlight relationships between points or events, as in Timemapper. Notably, the case studies revealed that even if relationships are demonstrated through lines, without the aid of highlighted routes, information can remain confusing to end-users.

It is also of note that all available programs reflect Western mapping principles and history, and little attention is paid to other cultural views. Ian Gregory (2010) and Ian Johnson (2011) both caution creators of digital maps to think carefully about what they choose to represent, as users frequently take maps as objective and without question. Creating maps is always a subjective process (Harpold, 1999).

Further Discussion

In spite of these limitations, the case studies suggest that visualizations do have the potential to add to the book history discipline. Although Palladio is a weaker tool, it showed how an aesthetic visualization can quickly highlight important historical hubs of book trade and movement. With even more data, more patterns may become visible, informing common trade and travel routes at certain times.

The case studies also suggest visualizations could play a role in addressing the challenges of provenance as metadata. Although metadata had to be modified for each tool studied, creating metadata for visualization helped make clear how the complexities of provenance metadata can be captured. This can provide stronger guidelines in the capture of provenance, and inform changing standards. As metadata needs to be reviewed in any case for use in visualization, an opportunity is also provided to deal with backlog.

Visualizations also help meet many of the contemporary uses of metadata. Almost all the tools examined provided an opportunity to create linked data, and through the cleaning of metadata, interoperability is greatly increased between collections, and data curation is accommodated. Although no case study offered the ability to search, some visualizations in the environmental scan did, suggesting searching visualizations is a feasible option to incorporate. Similarly, although issues of big data remain one of the greater limitations of the case studies, the use of filters as suggested by TimeMap could solve this problem, particularly if users select what they would like to see as an initial interaction with the tool. In general, both the environmental scan and the case studies demonstrate that many features are possible to offer in a tool that can enrich a user's experience, and offer them new ways of finding and understanding information.

No tool met all the criteria outlined perfectly, but of all the tools evaluated, Neatline is the one with the most features to display provenance with the richest nuances. Although it was not the easiest tool to use in the study, due to its requirement of basic coding skills, it is a good compromise between the heavy skills required by TimeMap and the light skills in Storymap JS, for the power of the visualization created. Further, many libraries already have an Omeka infrastructure, which allows relatively easy

implementation of additional plugins. Users can easily use different kinds of maps, and highlight relationships between points. Neatline also allows for ambiguities in data to be well represented. Notably, however, all the tools offered some advantage to users.

Special note should be made of TimeMap, as this tool is exceedingly flexible in what is portrayed and how – it was the only tool to include collections as a means of searching or sorting data, as well as time and place. The dated nature of the tool, however, combined with the fact that it requires the most specialized experience, prevent it from being the best option. In light of the many projects in development similar to TimeMap, however, one of those may be better than those evaluated here in the future.

Another special note should be made for Timemapper, primarily for its incredibly easy-to-use platform. Users do not require special training to use Timemapper, or experience in coding. Anyone with access can add events to the Google Spreadsheet, even well into the future, and no extra research is required to create location metadata. The aesthetic display with a large window for item description should also be praised. However, inability to display data ambiguity prevents it from being recommended.

Both Storymap JS and Palladio are the least favourable options. Although Storymap JS is aesthetic in its layout, only one journey is shown at a time. This precludes the tool from being suitable for the proposed use. Palladio offers an intriguing means of highlighting relationships via the use of darkened paths for an item, so that one journey stands out amongst many points and lines on the map. However, lack of additional information or means of navigation prevents Palladio from being a good option.

It is of note overall, however, that what remains the most important to creating a good visualization, is good and consistent metadata. No one tool accomplished

everything suggested in the criteria, although some come close. From the environmental scan and the case studies, however, we can see that the potential exists for a rich visualization that can meet criteria while also providing insight into metadata guidelines. Provenance metadata and visualization are on the precipice in digital collections of being a powerful Next Generation tool.

CHAPTER SEVEN: CONCLUSION AND RECOMMENDATIONS

At the outset of this thesis, a number of research questions were developed in order to investigate identified gaps in provenance metadata. In order to answer these research questions, historical research on provenance theory was conducted and a two-part action research study was designed and carried out.

The first part of the study was a comparative study of the use of provenance in sixty-four different digital special collections, from Canada, the United Kingdom and Ireland, and the United States. The standards used for the provenance metadata were also examined, as was the relation between select descriptive and digital features with provenance metadata, and the presence of the axes of provenance within that metadata. The second part of the study was a series of short case studies on five different visualization tools, which were analyzed to determine which visualization was the most productive in displaying provenance for end-users.

Provenance Metadata Standards

Historical research revealed that nearly forty-two different fields exist to capture provenance metadata, across an identified twenty-six standards. There is a plurality of reasons so many metadata fields exist across so many standards.

One of those reasons is the complexity of provenance itself as a theory and concept. While some metadata fields represent static information, such as publishers or titles, provenance itself refers to the origin or history of ownership of an item, a meaning that can include many kinds of nuanced information, and complex or missing pieces. In fact, at the very least, provenance refers to three identified axes of information: chains of owners, places, and time, although the ways these axes interact, physical marks of these

interactions, and the transfer between events are also part of provenance. This can make provenance difficult to restrain to a single metadata field.

Provenance metadata has also been largely shaped by its complex history with different communities and institutions, in particular, those of archives and special collections. Different communities have placed different emphases on provenance based on their needs. Provenance has long been a cornerstone of archival theory, although the interpretation of provenance has evolved over time from a record of ownership and means of organization towards including more social and cultural context.

The development provenance theory underwent in archives is important to understanding provenance within special collections libraries, which due to the nature of their unique materials are closer to archives than any other library institutions. Archives used free text *finding aids* to record provenance, and tended to do so as seen fit from institution to institution, likely influencing special collections to also develop their own local interpretations of provenance.

Provenance, however, never became a foundational principle in libraries as it did in archives. It developed as an official aspect to capture in special collection standards later, and always as an adjunct to overall guidelines. Combined with the development of varying terms, special collections captured provenance metadata inconsistently. This extended to the depth of information to capture, an issue identified in the literature (Pearson, 1994).

Metadata itself has many different types and formats that provenance can apply to, which may also affect how and what information is captured. With the development of digital environments, contemporary uses of metadata must also be addressed. Further,

standards change frequently over time, leading to increased complexity. This is likely to continue, as provenance theory is currently experiencing a surge of research in book history, leading only to increased interest and conflation of definitions of provenance across disciplines.

When coming to understand the multiple standards used to capture provenance metadata, it is also important to understand that metadata itself is also a complex entity, prone to errors from humans, or to restrictions of time and finance. No metadata schema is perfect, and data will always need revisiting and cleaning. As provenance is growing in digital fields, it is likely to only become more complex. The incentive of projects that seek to use metadata in new ways can therefore be a boon to help make metadata more robust in light of these limitations.

Current Use of Provenance Metadata

In "Provenance Metadata in Digital Special Collections", the first half of the research study, which examines the current use of provenance metadata in digital special collections, is presented. Findings revealed that compared to previous studies (see Overmier & Doak, 1996), provenance is being captured in collections at a higher rate than ever before. 80% of digital collections studied provided provenance data, and just under half of those offer at least three axes of provenance in that metadata. There is a definite increase in use and potential quality of provenance metadata. This is particularly highlighted by the use of the new element "provenance" by several collections, suggesting an identified need to provide such information.

In spite of this, at least twenty-five elements were used to capture provenance by these digital collections, with emphasis on the use of local or mixed elements, and

multiple standards were engaged. It seems provenance metadata is still in need of guidelines and structure.

When examined alongside descriptive factors, findings revealed that provenance metadata was more likely to be present and provide three axes in collections with overt ties to archives, in younger physical or digital collections, and in smaller collections. Types of institutions could play a role in the presence of provenance metadata, as museum institutions were the most likely to offer provenance metadata, although with only one axis of provenance; academic institutions were the second most likely to offer provenance, but with two or three axes of provenance; digital projects had a high likelihood to omit provenance, but tended between the extremes three or one axes; and national libraries were the least likely to offer provenance metadata, although when offered, it often had two or three axes of provenance.

Geographic locations of the host institutions could also play a role in the presence of provenance metadata. Canadian institutions all offered provenance, and collections were fairly evenly distributed according to the axes of provenance provided in the metadata. United Kingdom and Ireland institutions were the second most likely to provide provenance metadata but generally with two or three axes, and United States institutions were the least likely to offer provenance metadata, reflecting their unique histories with provenance theory.

When examining digital factors, findings suggested that where digital platforms are used, use of provenance metadata and the axes contained within that metadata is generally at the discretion of the institution, with the exception of the LUNA platform, where provenance was always provided. Use of linked metadata tended to occur in

collections with more location-based provenance information, suggesting that linked data is used in collections with more contextually-oriented metadata. At the same time, collections with less linked metadata were more likely to offer provenance with two or three axes, suggesting that it is easier to created linked metadata if there is less data to manage.

Metadata visualizations were identified, always in collections where there was provenance metadata. Those visualizations tended to be simple, using metaphors such as timelines, networks, and digital maps, without combining metaphors, and were generally based on provenance with only one axis, suggesting that current visualizations are not complex enough to support complex data.

Although the presence of provenance metadata and the number of axes provided in that metadata has improved in the past twenty years, many challenges remain in capturing provenance metadata. These include a lack of guidelines in capturing provenance metadata, dealing with changing standards, and managing large amounts of backlog in cataloguing data. User needs can therefore be adversely affected by the resulting influence on provenance metadata.

Visualizing Provenance

The second half of the action research study, presented in "Visualizing Provenance Metadata", explores the potential of provenance metadata visualization as a means of meeting the needs of users, while also taking advantage of the contemporary uses of metadata as a Next Generation catalogue tool for special collection libraries. These uses include linked data, interoperability, federated searches, sorting big data, and data curation.

An environmental scan of eight digital projects which combine visualization metaphors of place and time, as well as case studies of five open-source, geotemporal tools demonstrated the numerous features tools can make available to end-users in order to help highlight facets of provenance, including ownership, time, place, and material.

Of the five case studies, the tool Neatline was determined to be the best option for collections seeking to visualize the provenance of their items, as it provided the best ability to demonstrate nuance and complexities inherent in provenance metadata for the time and skill investment necessary. No tool evaluated, however, offered all the desirable features suggested from the literature and the environmental scan criteria, in particular a means of searching and sorting big data. The study did suggest, however, that visualizations do offer the potential to meet user needs and help work towards solutions for many of the challenges of provenance metadata, by helping to standardize metadata and provide opportunities for libraries to address backlog in metadata.

Beyond the conclusions of the two-part action research study presented in this thesis, a number of recommendations can also be extrapolated for both the capture of provenance metadata and its visualization.

Recommendations

The complex history and nature of provenance theory informed the standards developed to capture that information. These standards and guidelines, in turn, have affected what is currently available for users in digital environments. Results from this thesis suggest that provenance metadata is currently benefitting from a height in interest and research into provenance theory, and that provenance metadata itself is the best quantity and quality it has ever been in. Further, current visualization instruments for

provenance metadata are on the cusp of being excellent Next Generation catalogue tools. Both provenance metadata and visualization tools, however, could benefit from recommendations to help special collection catalogues in their designs for end-users.

Metadata recommendations

As a substitute for examining a book, copy-specific metadata plays a particularly important role. It allows researchers to work at a distance from material, supports new modes of research, and "promot[es] preservation of the item itself" (Lundy, 2008, p. 166). In spite of this, metadata can be difficult to work with due to the plethora of standards available, and the fact that it will always require work and human intervention to be useable. In light of this fact, maintaining good guidelines and quality is key to meeting user needs and encouraging interaction and growth of online catalogues.

As outlined above, provenance metadata is particularly complex, and in fact is multi-faceted (Pearson, 1994; see "Provenance Theory"). Further, provenance can be even more complicated as each axis has the potential to interact with other axes in many-to-many relationships. All of this evidence suggests that provenance stretches across many types of metadata, and in fact should not be one element alone. In other words, provenance is actually comprised of several potential fields.

This is made particularly apparent when preparing provenance metadata for visualization, where information must be broken up in several pieces to be understood by the tool. Extrapolation from the findings in this thesis as well as suggestions within the literature (Pearson, 1994) can be used to create a set of recommendations for capturing provenance metadata as multiple fields (see *Table 6*).

These fields can be used to capture all the facets of provenance, such as owners, place, and time, as well as the means of transference. Copy-specific information and evidence of provenance can also be readily identified, as recommended by Pearson (1994) through the use of inscription, bookplate, and armorial stamp fields. Finally, each step in the chain of evidence is clearly laid out, as recommended by Pearson (1994). Provenance metadata in this format lends itself well to visualization, as the information becomes both more standard, and of better quality.

Table 6

Recommended Data Fields for Provenance Information

Title	
Author	
Date of Publication	
Location of Publication	
Inscription	
Bookplate	
Armorial Stamp	
Item Description	

Owner 1		Provenance Date(s) 1		Provenance Location 1		Description of ownership event 1	
Owner 2		Provenance Date(s) 2		Provenance Location 2		Description of ownership event 2	
Owner 3		Provenance Date(s) 3		Provenance Location 3		Etc.	

Table 6

Visualization Recommendations

No tool examined in the study made use of all the identified features for visualizing provenance metadata effectively for users; however, the case studies and environmental scan helped demonstrate what is possible to create.

Creating a new geotemporal tool, while a lot of work, would be possible, and potentially worthwhile in view of the larger trends libraries are headed towards with Next Generation catalogues, international provenance databases, and meeting user needs. Developing a new tool specifically for provenance metadata would allow all three axes to be represented visually for multiple entities, thereby accounting for shortcomings of existing software, while simultaneously providing a goal and guide for cleaning and enriching provenance metadata. It would also allow the challenges of provenance metadata to be addressed directly.

Figure 28 presents a mock-up of what such a tool might look like. Features included in this mockup include not only the use of timelines and digital maps, but the ability to select different collections, items, or owners to see in the visualization, as seen in the TimeMap tool, to allow users to sort through large amounts of data. This information can also be colour-coded, such as in TimeMap.

The digital map itself would ideally be customizable by end-users, such as through choosing to switch to an antiquated map, as Neatline allows users to do, as well as enabling the ability to zoom in and out in order to see clusters more clearly. Use of a mini-map to navigate may also be helpful, as featured by Storymap JS and Hypercities.

Another important feature that would be of use to end-users is conveying a strong sense of chronology, such as accomplished by Storymap JS or Palladio, allowing users to

see related events through highlighted lines. End-users should also be able to search information, as the Heurist tool or the Bookbinding Map allow. The ability to show ambiguity in location and date is also necessary in a good tool. For example, Neatline allows the selection of any size of geographic area, or the inclusion of dates or places without corresponding place or date information respectively.

Additional information that may inform the visualization should be available through generous room given to read about a selected item, as demonstrated by the Timemapper tool. This way, links, images, descriptions of modes of transfer or of material, as well as information about bookbindings or inscriptions can be provided. Pop-up information for selected points would also be ideally provided, as well as the ability to share information through social media. Finally, collections should also be able to easily contribute data in the future, such as through Timemapper's Google Spreadsheet.

Inclusion of these described features would cover all the facets outlined related to provenance: space, time, personality (ownership), matter (physical material), and energy (method item changed hands), as well as many of the criteria identified through the literature and environmental scan.

Figure 28: Mockup of a potential visualization of provenance metadata for library catalogues

Such a prototype would require further refinement, but the case studies and environmental scan help inform its possibilities. It also provides multiple points of access to information for users, while breaking down provenance into step-by-step chronologies, as recommended by Pearson (1994). After development of a prototype, a large sample of data should be tested, and a short usability study of humanist scholars performed in order to evaluate the interface.

Future research

In order to build upon the recommendations suggested by this thesis, a visualization prototype remains to be developed, tested, and integrated into a library

catalogue. A successful visualization prototype will provide an opportunity for catalogues to use new digital tools, with increased meaning for users.

The success of integrating visualizations of provenance metadata opens the door to many more research projects for libraries. The development of a "take-a-trip" feature would be useful for users, allowing the journey of a single book to be displayed through video-like scenario. Users may also wish to save their searches and browses. Collections may want to consider developing provenance visualizations for a mobile app. As digital tools develop, libraries will have to be conscious of their user base, and strive to incorporate different cultural concepts of time and space to create meaningful tools. Libraries should also strive to use accessible technology.

Widespread application of digital mapping and multimedia overlay of rare books and manuscripts also provides opportunity for crowd-sourcing information about the provenance of an item (see Shaw, 2012). By encouraging user participation through digital tools, cataloguers may gain insight from experts in the field, which helps create the most accurate information possible. Such practices can save time, finances, and research for libraries in creating richer provenance metadata, and aligns with the current interests of archivists in involving communities in the capture of provenance metadata.

A developed provenance visualization holds application beyond rare book libraries: it may also be of use in museums and archival practice. Provenance metadata is an important field to consider in an increasingly digital world as well, as born digital items have complex relationships with different print mediums and digital environments. Capturing this provenance information is becoming increasingly important, especially in preserving information in a world where technology changes quickly. Examining and

contrasting tools built for born digital items with tools used for rare book provenance may be a fruitful study illuminating potential advantages for both fields. Finally, visualizations can map more than just the journeys of rare books – it may be of interest to trace the journeys of modern library books, to see what branches in a public library a book is most frequently loaned to, for example.

In an increasingly digital world, provenance metadata has grown in importance for conveying vital information to scholars who may not have easy access to a physical version of a rare book or manuscript. Because of this, stricter metadata rules need to be enforced. However, digital tools have also opened up a whole new realm of possibility for libraries to reach their users through, enhancing and enriching current catalogues and systems. In light of changing library standards, and the goals of Next Generation catalogues, provenance visualization provides a rich opportunity for libraries to present their users with not only new means of discovering information, but with new tools to understand and further their disciplines as well.

REFERENCES

"provenance, n." (2014, December). *OED Online*. Oxford University Press. Retrieved from http://www.oed.com.login.ezproxy.library.ualberta.ca/view/Entry/153408?redirectedFrom=provenance

Adam Matthew Digital. (n.d.a). Literary Manuscripts Leeds. Retrieved from http://www.amdigital.co.uk/m-collections/collection/literary-manuscripts-leeds/

Adam Matthew Digital. (n.d.b). Medieval Travel Writing. Retrieved from www.medievaltravel.amdigital.co.uk

Adam Matthew Digital. (n.d.c). Perdita Manuscripts, 1500-1700. Retrieved from http://www.amdigital.co.uk/m-collections/collection/perdita-manuscripts-1500-1700/

Adams, F. B. (1969). *The Uses of Provenance*. Los Angeles: University of California.

admin. (2015, October 20). Google Location History Now Shows Timeline [Web log entry]. Retrieved from http://techkeeg.com/2015/10/20/google-location-history-now-shows-timeline/

Agricultural Information Management Standards (AIMS). (2014). National Library of Medicine (NLM) Metadata Schema. Retrieved from http://aims.fao.org/fr/vest-registry/metadata-sets/national-library-medicine-nlm-metadata-schema

American Museum of Natural History. (n.d.). Accessioning. Retrieved from http://collections.paleo.amnh.org/4/accessioning

Anagnostopoulos, I., Zeadally, S., & Exposito, E. (2016). Handling big data: research challenges and future directions. *The Journal of Supercomputing, 72*(4), 1494-1516. Doi:10.1007/s11227-016-1677-z.

Annotated Books Online. (n.d.a). Annotated Books Online. Retrieved from http://abo.annotatedbooksonline.com/

Annotated Books Online. (n.d.b). Background. Retrieved from http://www.annotatedbooksonline.com/background/

Araya, A. (2003). The Hidden Side of Visualization. *Virginia Techné, 11* (1).

Baca, M., Clarke, S., Eklund, J., Gilliland, A. J., Harpring, P., Woodley, M. S., & O'Keefe, E. (2014). Metadata Standards Crosswalk. Retrieved from http://www.getty.edu/research/publications/electronic_publications/intrometadata/crosswalks.html

Bartlett, N. (1992). 'Respect des fonds:' The origins of the modern archival principle of provenance. *Primary Sources & Original Works, 1*(1/2), 107-115.

Bayih, Y. T. (2010). Application of Preservation Metadata for Long-Term Accessibility of Digital Objects. In A. Katsirikou, & C. Skiadas. (Eds.), *New Trends in Qualitative and Quantitative Methods in Libraries: Selected Papers Presented at the 2nd Qualitative and Quantitative Methods in Libraries* (167-173). New Jersey: World Scientific.

Beal, V. (2015). Big data. Retrieved from http://www.webopedia.com/TERM/B/big_data.html

Bearman, D. A., & Lytle, R. H. (1985). The Power of the Principle of Provenance. *Archivaria, 21*, 14 – 27.

Bergström, P., & Atkinson, D. C. (2010). Augmenting Digital Libraries with Web-Based Visualizations. *Journal of Digital Information Management, 8*(6), 377-386.

Binding, C., & Tudhope, D. (2016). Improving interoperability using vocabulary linked data. *International Journal on Digital Libraries*, *17*(1), 5-21.

Bodleian Libraries University of Oxford. (2015). About Us. Retrieved from http://www.bodleian.ox.ac.uk/about-us

Bodleian Library. (2015). Digital Bodleian. Retrieved from http://digital.bodleian.ox.ac.uk/

Borbinha, J., Pedrosa, G., Gil, J., Martins, B., Freire, N., Dobreva, M., Wyttenbach, A. (2007). Digital libraries and digitised maps: an early overview of the DIGIMAP project. In D. H. Goh, T. H. Cao, I. T. Sølvberg,& E. Rasmussen (Eds.), *Asian Digital Libraries. Looking Back 10 Years and Forging New Frontiers: 10th International Conference on Asian Digital Libraries, ICADL 2007, Hanoi, Vietnam, December 10-13, 2007. Proceedings* (383-386). Berlin: Springer Berlin Heidelberg.

Botanicus. (n.d.a). About Botanicus. Retrieved from http://www.botanicus.org/About.aspx

Botanicus. (n.d.b). Botanicus Digital Library. Retrieved from http://www.botanicus.org/browse

Brashear, R. (2004). Introduction: Incunabula in the Dibner Library of the History of Science and Technology. Retrieved from http://www.sil.si.edu/digitalcollections/incunabula/intro.htm

British Library. (n.d.a). Digitised Manuscripts. Retrieved from http://www.bl.uk/manuscripts/

British Library. (n.d.b). English Short Title Catalogue – history. Retrieved from http://www.bl.uk/reshelp/findhelprestype/catblhold/estchistory/estchistory.html

British Library. (n.d.c). Timelines: Sources from History. Retrieved from http://www.bl.uk/learning/histcitizen/timeline/accessvers/index.html

British Library. (n.d.d). Treasures in Full: High-quality digital editions, free to your desktop. Retrieved from http://www.bl.uk/treasures/treasuresinfull.html

British Library. (2016). Timelines: Sources from History. Retrieved from http://www.bl.uk/learning/timeline/index.html

British Library Board. (n.d.). English Short Title Catalogue. Retrieved from http://estc.bl.uk/F/?func=file&file_name=login-bl-estc

Brown, S. & Considine, J. (2010). *Marginated: Seventeenth-Century Printed Books and the Traces of Their Readers*. Edmonton: University of Alberta Libraries.

Bruni, F. (2011). From Inventories to Signs on Books: Evidence for the History of Libraries in the Modern Age. *The International Journal of the Book, 8*(4), 51-61.

Buchanan, S. A. (2011). Special Collections and Archives Code (SPAC): Preserving Provenance in a Bibliographic Catalog. *Cataloging & Classification Quarterly, 49*(5): 349-360. doi: 10.1080/01639374.2011.580446

Bureau of Canadian Archivists. (2008). *Rules for Archival Description*. Ottawa: Canadian Cataloguing in Publication Data. Retrieved from http://www.cdncouncilarchives.ca/rad/radcomplete_july2008.pdf

Burke, M. (2009). The semantic web and the digital library. *Aslib Proceedings, 61*(3), 316-322.

Cambridge Digital Library. (2015a). Retrieved from http://cudl.lib.cam.ac.uk

Cambridge Digital Library. (2015b). Introducing the Cambridge Digital Library. Retrieved from http://cudl.lib.cam.ac.uk/

Carless, J. (2014). Collecting and authenticating online evidence: Now more than ever, proving chain of custody matters. Retrieved from http://www.cba.org/cba/practicelink/careerbuilders_technology/web-evidence.aspx

Carter, J., & Barker, N. (2004). *ABC for Book Collectors* (Eighth Edition). New Castle: Oak Knoll Press.

Caswell, J. V., & Wynstra, J. D. (2010). Improving the search experience: federated search and the library gateway. *Library Hi Tech, 28*(3), 391-401.

Challenges and Issues with Metadata Crosswalks. (2002). *Online Libraries & Microcomputers, 20*(4), 1-4.

Chartered Institute of Library and Information Professionals (Great, Britain), Joint Steering Committee for Development of RDA, American Library Association, & Canadian Library Association. (2013). *RDA: Resource Description & Access*. Chicago, Illinois: American Library Association.

Chen, M., Mao, S., Zhang, Y., & Leung, V. C. M. (2014). *Big Data: Related Technologies, Challenges, and Future Prospects*. New York: Springer International Publishing.

Chickering, F. W., & Yang, S. Q. (2014). Evaluation and Comparison of Discovery Tools: An Update. *Information Technology & Libraries, 33*(2), 5-30.

Chowdhury, G.G., & Chowdhury, S. (2000). An overview of the information retrieval features of twenty digital libraries. *Program, 34*(4). 341-373.

Christie's. (2009, April 20). Case History: The Importance of Provenance. Retrieved from http://www.christies.com/features/2009-April-Importance-of-Provenance-56-1.aspx

Columbia University Libraries / Information Services. (n.d.). Rare Books & Manuscript Library. Retrieved from http://library.columbia.edu/locations/rbml.html

Consortium of European Research Libraries (CERL). (2014). Provenance Information. Retrieved from http://www.cerl.org/resources/provenance/main

Consortium of European Research Libraries (CERL). (n.d.a). Caxton, William. *CERL Thesaurus*. Retrieved from http://thesaurus.cerl.org/cgi-bin/record.pl?rid=cni00036072

Consortium of European Research Libraries (CERL). (n.d.b). CERL Resources. Retrieved from http://www.cerl.org/resources/main

Consortium of European Research Libraries (CERL). (n.d.c). Searching for Provenance Information. Retrieved from http://www.cerl.org/resources/provenance/main

Cook, T. (1993). The Concept of the Archival Fonds in the Post-Custodial Era: Theory, Problems, and Solutions. *Archivaria, 35*, 24 – 37.

Cook, T. (1997). What is Past is Prologue: A History of Archival Ideas Since 1898, and the Future Paradigm Shift. *Archivaria, 43*, 17- 63.

Cook, T. (2013). Evidence, memory, identity, and community: four shifting archival paradigms. *Archival Science, 13*: 95-120. doi: 10.1007/s10502-012-9180-7

Council on Library and Information Resources. (2014). Data Curation. Retrieved from http://www.clir.org/initiatives-partnerships/data-curation

CRS. (1991). Descriptive Cataloging of Rare Books, 2nd Edition: Contents. Retrieved

from http://www.itsmarc.com/crs/mergedProjects/rarebks/rarebks/contents.htm

Curwen, T., & Jonsson, G. (2007). Provenance and the Itinerary of the book: recording provenance data in on-line catalogues. In D. J. Shaw (Ed.). *Imprints and owners: recording the cultural geography of Europe: papers presented on 10 November 2006 at the CERL Seminar hosted by the National Szechenyi Library, Budapest*. (31-48). London: CERL, 2007.

Davis, A. P. (2014). Provenance tracking. Retrieved from http://rrcns.readthedocs.org/en/latest/provenance_tracking.html#references

Davison, S. (2009). If We Build It, Will They Come? Strategies for Teaching and Research with Digital Special Collections. *RBM: A Journal of Rare Books, Manuscripts, & Cultural Heritage, 10*(1), 37-49.

DCC. (2014). What are Metadata Standards. Retrieved from http://www.dcc.ac.uk/resources/briefing-papers/standards-watch-papers/what-are-metadata-standards

DCC. (2015). CONTENTdm. Retrieved from http://www.dcc.ac.uk/resources/external/contentdm

DCMI. (2002). Dublin Core Collective Description Proposed Term : Provenance. Retrieved from http://www.ukoln.ac.uk/metadata/dcmi/collection-provenance/

Deckelbaum, D. (1999). GIS in Libraries: An Overview of Concepts and Concerns. *Issues in Science and Technology Librarianship*. Retrieved from http://webdoc.sub.gwdg.de/edoc/aw/ucsb/istl/99-winter/article3.html

Del Rio, N. & da Silva, P. P. (2007). Probe-It! Visualization Support for Provenance. In D. Hutchinson, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, &

N. Paragios (Eds.), *Advances in Visual Computing* (732-741). Berlin: Springer Berlin Heidelberg. doi:10.1007/978-3-540-76856-2_72

Derrida, J. (1995). Archive Fever: A Freudian Impression (E. Prenowitz, Trans.). *Diacritics, 25*(2), 9-63.

Diem, A. (2011). Monastic Manuscript Project. Retrieved from http://www.earlymedievalmonasticism.org/about.html.

Digital Image Archive of Medieval Music. (2012). Welcome to DIAMM. Retrieved from http://www.diamm.ac.uk/

Digital Scriptorium. (n.d.a). About the Digital Scriptorium. Retrieved from http://bancroft.berkeley.edu/digitalscriptorium/about/

Digital Scriptorium. (n.d.b). Digital Scriptorium. Retrieved from http://bancroft.berkeley.edu/digitalscriptorium/

Dooley, J. M., Beckett, R, Cullingford, A., Sambrook, K., Sheppard, C., & Worrall, S. (2013). Survey of Special Collections and Archives in the United Kingdom and Ireland. Dublin, Ohio: OCLC Online Computer Library Center, Inc. Retrieved from http://www.oclc.org/content/dam/research/publications/library/2013/2013-01.pdf

Douglas, J. (2010). Origins: Evolving Ideas about the Principle of Provenance. In T. Eastwood, & H. MacNeil (Eds.). *Currents of archival thinking*. (23-43). Santa Barbara: Libraries Unlimited.

Drupal. (n.d.). Time Map. Retrieved from https://www.drupal.org/project/timemap

Duchein, M. (1992). The history of European archives and the development of the archival profession in Europe. *American Archivist, 55*, 14-25.

Earhart, A. E. (2014). 'After a hundred years/Nobody knows the Place': Notes Toward Spatial Visualizations of Emily Dickinson. *Emily Dickinson Journal, 23*(1), 98-105.

Early English Books Online. (2015). Early English Books Online. Retrieved from http://eebo.chadwyck.com/home

Echard, S. (n.d.). Medieval Manuscripts on the Web. Retrieved from http://faculty.arts.ubc.ca/sechard/512digms.htm

Egenhofer, M. J. (2002). Toward the Semantic Geospatial Web. In *Proceedings of the 10th ACM international symposium on Advances in geographic information systems (GIS'02)* (1-4). New York: ACM.

Elings, M. W., & Waibel, G. (2007). Metadata for all: Descriptive Standards and Metadata Sharing across Libraries, Archives, and Museums. *First Monday, 12* (3). Retrieved from http://firstmonday.org/article/view/1628/1543

European Library. (2013). The European Library: Connecting knowledge. Retrieved from http://www.theeuropeanlibrary.org/tel4/

Fagan, J. C. (2010). Usability Studies of Faceted Browsing: A Literature Review. *Information Technology & Libraries, 29*(2), 58-66. Retrieved from http://commons.lib.jmu.edu/cgi/viewcontent.cgi?article=1049&context=letfspubs

Farman, J. (2012). Chapter 2: Mapping and Representations of Space. In *Mobile Interface Theory* (35-55). London: Routledge.

Fitzwilliam Museum. (2015). Collections Explorer. Retrieved from http://webapps.fitzmuseum.cam.ac.uk/explorer/

Folger Shakespeare Library. (n.d.). Browse Folger Digital Image Collection. Retrieved from http://luna.folger.edu/luna/servlet/FOLGERCM1~6~6

From DCRB to DCRM(B). (n.d.). Retrieved from http://www.cilip.org.uk/sites/default/files/documents/DCRBvsDCRMB.pdf

Gbdirect. (n.d.). Benefits of Using Open Source Software. Retrieved from http://open-source.gbdirect.co.uk/migration/benefit.html

Gill, T. (2004). Building semantic bridges between museums, libraries, and archives: The CIDOC Conceptual Reference Model. *First Monday, 9*(5). Retrieved from http://firstmonday.org/ojs/index.php/fm/article/viewArticle/1145

Gingerich, O. (2005). *The Book Nobody Read: Chasing the Revolutions of Nicolaus Copernicus*. New York: Walker Publishing Company.

GitHub. (2015). Cwrc/CWRC-Mapping-Timelines-Project. Retrieved from https://github.com/cwrc/CWRC-Mapping-Timelines-Project

Göbel, S., & Jasnoch, U. (2001). Visualization techniques in metadata information systems for geospatial data. In *Advances in Environmental Research: International Symposium on Environmental Software Systems (ISESS 2000), 5*(4), 415-424.

Google Code. (n.d.). timemap. Retrieved from https://code.google.com/archive/p/timemap/

Google Maps GPS Coordinates. (n.d.). Retrieved from http://www.gps-coordinates.net/

Gorman, G. E. & Clayton, P. (2005). *Qualitative research for the information professional: A practical handbook, Second edition*. London: Facet publishing.

Greenberg, J. (2005). Understanding Metadata and Metadata Schemes. *Cataloging & Classification Quarterly, 40* (3/4), 17-36.

Greene, S., Marchionini, G., Plaisant, C., Shneiderman, B. (2000). Previews and overviews in digital libraries: Designing surrogates to support visual information

seeking. *Journal for the American Society for Information Science, 51*(4), 380-393.

Gregory, I. (2010). Exploiting Time and Space: A Challenge for GIS in the Digital Humanities. In D. Bodenhamer, J. Corrigan, & T. M. Harris. (Eds.). *The Spatial Humanities: GIS and the Future of Humanities Scholarship* (58-75), Bloomington: Indiana University Press.

Groth, P. (2007). *The Origin of Data: Enabling the Determination of Provenance in Multi-institutional Scientific Systems through the Documentation of Processes* [Doctoral Dissertation]. University of Southampton, School of Electronics and Computer Science.

Hallo, M., Luján-Mora, S., Maté, A., & Trujillo, J. (2016). Current state of Linked Data in digital libraries. *Journal of Information Science, 42*(2), 117-127.

Han, M. J., Cho, C., Cole, T. W., & Jackson, A. S. (2009). Metadata for Special Collections in Contentdm: How to Improve Interoperability of Unique Fields Through OAI-PMH. *Journal of Library Metadata, 9*(3/4), 213-238.

Harpold, T. (1999). Dark Continents: A Critique of Internet Metageographies. *Postmodern Culture, 9*(2). Retrieved from http://muse.jhu.edu.login.ezproxy.library.ualberta.ca/journals/postmodern_culture/v009/9.2harpold.html

Harpring, P. (2009). Cataloging in the Museum: CCO, CDWA, and CDWA Lite for Cataloging Museum Objects [Powerpoint Slides]. Retrieved from http://www.getty.edu/research/tools/vocabularies/cco_cdwa_for_museums.pdf

Harpring, P. (2014). *Categories for the Description of Works of Art (CDWA): List of Categories and Definitions*. Paul Getty Trust & College Art Association, Inc.

Retrieved from http://www.getty.edu/research/publications/electronic_publications/cdwa/definitions.pdf

Harry Ransom Center, The University of Texas at Austin. (n.d.). Medieval and Early Modern Manuscripts Collection: Database and Digital Images. Retrieved from http://norman.hrc.utexas.edu/pubMnEM/

Hartig, O. (2009, April). Provenance Information in the Web of Data. In *Proceedings of the Linked Data on the Web Workshop (LDOW'09)*, Madrid, Spain. Retrieved from http://events.linkeddata.org/ldow2009/papers/ldow2009_paper18.pdf

Harvard College Library. (2008). Houghton Library: History. Retrieved from http://hcl.harvard.edu/libraries/houghton/history.cfm

Harvard College Library. (2015). Houghton Library. Retrieved from http://hcl.harvard.edu/libraries/houghton/

Heath, T. (n.d.). Linked Data. Retrieved from http://linkeddata.org/home

Hedstrom, M., & King, J. L. (2003). *On the LAM: Library, archive, and museum collections in the creation and maintenance of knowledge communities*. Paris, France: Organisation for Economic Co-operation and Development. Retrieved from http://www.oecd.org/dataoecd/59/63/32126054.pdf

Hensen, S. L. (1989). Archives, personal papers, and manuscripts: a cataloguing manual for archival repositories, historical societies, and manuscript libraries. Chicago: Society of American Archivists. Retrieved from http://catalog.hathitrust.org/Record/002619326

Hensen, S. L. (1993). The First Shall Be First: APPM and Its Impact on American

Archival Description. *Archivaria 35*, 64-70.

Heurist. (n.d.). Retrieved from http://heuristnetwork.org/

Hill, L. L., Janée, G., Dolin, R., Frew, J., & Larsgaard, M. (1999). Collection Metadata Solutions for Digital Library Applications. *Journal Of The American Society for Information Science, 50*(13), 1169-81.

Hillmann, D. I. (2008). Cataloging for the 21st Century – Course 2. Metadata Standards & Applications: Trainee Manual. [Powerpoint Slides]. Retrieved from http://loc.gov/catworkshop/courses/metadatastandards/pdf/MSTraineeManual.pdf

Hillyard, B., & Pearson, D. (2007). *Guidelines for the Cataloguing of Rare Books*. London: CILIP Rare Books and Special Collections Group. Retrieved from http://www.cilip.org.uk/sites/default/files/documents/Guidlines%20for%20the%20Cataloguing%20of%20Rare%20Books.pdf

Hirsch, R. (2010). The Permanence of Provenance: The "Two Traditions" and the American Archival Profession. *Journal of Archival Organization, 8*(54), 54-72. doi:10.1080/15332748.2010.486754

History of Information. (n.d.). Bookbinding Map. Retrieved from http://www.historyofinformation.com/maps-simple/?category=Bookbinding&era=&region=

Howarth, L. C., Moor, L., & Sze, E. (2010). Mountains to Molehills: The Past, Present, and Future of Cataloging Backlogs. *Cataloging & Classificaation Quarterly, 48*(5), 423-444.

Humanities+Design. (n.d.a). Open Access Tools. Retrieved from http://hdlab.stanford.edu/tools/

Humanities+Design. (n.d.b). Palladio. Retrieved from http://palladio.designhumanities.org/#/

Huntington Digital Library. (n.d.) All Collections. Retrieved from http://hdl.huntington.org/

Huntington. (2015). About the Hungtington. Retrieved from http://huntington.org/WebAssets/Templates/content.aspx?id=56

Hurley, C. (2005). Parallel Provenance: (1) What, if Anything, is Archival Description?. *Archives & Manuscripts, 33*(1), 110-145.

Hutton, J. (2008). Academic Libraries as Digital Gateways: Linking Students to the Burgeoning Wealth of Open Online Collections. *Journal Of Library Administration, 48*(3/4), 495-507.

ICOM. (2010). What is LIDO. Retrieved from http://network.icom.museum/cidoc/working-groups/data-harvesting-and-interchange/what-is-lido/

Ingersoll, R. V. (2014). Provenance (geology). In *AccessScience*. McGraw-Hill Education. Retrieved from http://www.accessscience.com/content/provenance-geology/757524

International Council of Museums. (2014). The CIDOC Conceptual Reference Model. Retrieved from http://www.cidoc-crm.org/

International Council on Archives. (2000). *ISAD(G): General International Standard Archival Description, Second Edition*. Ottawa: Canadian Cataloguing in Publication Data. Retrieved from http://www.icacds.org.uk/eng/ISAD%28G%29.pdf

International Foundation for Art Research (IFAR). (n.d.). Provenance Guide. Retrieved

from http://www.ifar.org/provenance_guide.php

Irish Script on Screen. (n.d.). Irish Script on Screen Home. Retrieved from https://www.isos.dias.ie/english/index2.html

J. Paul Getty Museum. (n.d.). Collection. Retrieved from http://www.getty.edu/art/collection/

Jacobsen, J. (2010, April 28). Encoded Archival Description (EAD) Export for DB/TextWorks. *Andornot Blog*. Retrieved from http://www.andornot.com/blog/post/Encoded-Archival-Description-%28EAD%29-Export-for-DBTextWorks.aspx

Johnson, I. (2011). Spatiality and the Social Web: Resituating Authoritative Content. In M. Dear, J. Ketchum, S. Luria, & D. Richardson (Eds.), *Geohumanities: Art, History, Text at the Edge of Place* (267-276). New Brunswick: Routledge.

Katsirikou, A., & Skiadas, C. (Eds.). (2010). *New Trends in Qualitative and Quantitative Methods in Libraries: Selected Papers Presented at the 2nd Qualitative and Quantitative Methods in Libraries*. Singapore: World Scientific.

Kennedy, P. (2008). Manifestations of metadata: from Alexandria to the Web – old is new again. *The Australian Library Journal*, *57*(2), 128-146.

Kumar, R. (2011). *Research Methodology: a step-by-step guide for beginners, Third Edition*. Los Angeles: Sage Publications.

Kumar, V. (2008). Next Generation Catalogue: A User's expectation. *Inflibnet centre*. Retrieved from http://hdl.handle.net/1944/1283

Kumar, V., Furuta, R., & Allen, R. B. (1998). Metadata Visualization for digital libraries: interactive timeline editing and review. In I. Witten, R. Akscyn, & F. M.

Shipman, III (Eds.), *DL '98 Proceedings of the third ACM conference on Digital Libraries* (126-133). New York: ACM.

Leslie, D. J. (1997). "Provenance Evidence" and "Printing and Publishing Evidence": Use and Revision of the RBMS Thesauri. *The Papers of the Bibliographical Society of America, 91*(4), 517-523.

Liber calculationis ('Salzburg encyclopedia of 818'). (n.d.). The European Library. Retrieved from http://www.theeuropeanlibrary.org/tel4/records/2000061252746

Library of Congress. (2007). *Descriptive Cataloguing of Rare Materials (Books)*. Washington, D.C.: Library of Congress.

Library of Congress. (2010). MODS <note> Types. Retrieved from http://www.loc.gov/standards/mods/mods-notes.html

Library of Congress. (2013). Medieval and Renaissance Digital Materials from the Lessing J. Rosenwald Collection. Retrieved from http://www.loc.gov/rr/rarebook/rosenwald-MandR.html

Library of Congress. (2014). METS Schema Documentation. Retrieved from http://www.loc.gov/standards/mets/docs/mets.v1-9.html

Library of Congress. (2015a). Rare Book & Special Collections Reading Room. Retrieved from http://www.loc.gov/rr/rarebook/

Library of Congress. (2015b). Rare Book & Special Collections Reading Room: Digitized Materials from the Rare Book & Special Collections Division. Retrieved from http://www.loc.gov/rr/rarebook/digitalcoll.html

Library of Congress. (n.d.a). Collection: Pre-1700 Musical Treasures: Manuscript and Print Collection. Retrieved from http://www.loc.gov/collections/pre-1700-musical-

treasures/

Library of Congress. (n.d.b). Music Treasures Consortium. Retrieved from http://lcweb2.loc.gov/diglib/ihas/html/treasures/treasures-home.html

Library of Congress. (n.d.c.). got metdata? [pdf]. Retrieved from http://core.vraweb.org/pdfs/core4_xml.pdf

Lichfield Cathedral. (2015). Manuscripts of Lichfield Cathedral. Retrieved from https://lichfield.as.uky.edu/

Lichfield Cathedral. (n.d.). Cathedral Treasures. Retrieved from http://www.lichfield-cathedral.org/treasures

LIDO- Lightweight Information Describing Objects. (2010). XML Schema for Contributing Content to Cultural Heritage Repositories Version 0.9. Retrieved from http://www.lido-schema.org/schema/v0.9/lido-v0.9-schema-listing.html

LIDO. (2010). What is LIDO. Retrieved from http://network.icom.museum/cidoc/working-groups/data-harvesting-and-interchange/what-is-lido/

Liu, J. (2007). *Metadata and its applications in the digital library*. Westport, CT: Libraries Unlimited.

Liu, W., Xia, C., Zhao, L., Lou, X., & Zhang, C. (2004). The Design of Metadata Schemas for Chinese Cultural Celebrities' Manuscript Library. *DC-2004—Shanghai Proceedings*. Retrieved from http://dcpapers.dublincore.org/pubs/article/view/785

LUNA Library. (n.d.). Retrieved from http://doc.lunaimaging.com/display/V7D/LUNA+Library

LUNA. (n.d.). Working with Data Templates. Retrieved from http://doc.lunaimaging.com/display/V7D/Working+with+Data+Templates

Lundy, M. W. (2008). Provenance Evidence in Bibliographic Records: Demonstrating the Value of Best Practices in Special Collections Cataloging. *Library Resources & Technical Services, 52*(3), 164-172.

Mackenzie, N., & Knipe, S. (2006). Research dilemmas: Paradigms, methods, and methodology. *Issues in Educational Research, 16*(2), 193-205. Retrieved from http://www.iier.org.au/iier16/mackenzie.html

Mantegari, G., & Edlestein, D. (2012). Voltaire's places of publication (1712-1800). Retrieved from http://republicofletters.stanford.edu/casestudies/voltairepub.html

Manuscripts and Archives Division, The New York Public Library. (2015). *Portrait of Matthew in gold frame on purple background, with gold, green, and other colors*. Retrieved from http://digitalcollections.nypl.org/items/510d47da-e296-a3d9-e040-e00a18064a99

Marquez, J. J. (2012). Extending the Library's Web Presence: Building an Online Timeline Using MIT's SIMILE Exhibit. *Journal of Library Innovation, 3*(2), 63-79.

Material Evidence in Incunabula (MEI). (n.d.). About this database. Consortium of European Research Libraries. Retrieved from http://incunabula.cerl.org/cgi-bin/search.pl

Mazza, R. (2009). *Introduction to information visualization*. London: Springer.

McCarty, W. (2005). *Humanities Computing*. New York: Palgrave Macmillan.

McGuinness, D. L., Ding, L., Glass, A., Chang, C., Zeng, H., & Furtado, V. (2006). Explanation Interfaces for the Semantic Web: Issues and Models. *Workshop on*

Semantic Web User Interactions, International Semantic Web Conference. Retrieved from http://ksl.stanford.edu/pub/KSL_Reports/KSL-06-14.pdf

McQuillen, J. T. (2012). *In Manuscript and Print: The Fifteenth-Century Library of Scheyern Abbey* [Doctoral Dissertation]. University of Toronto, Department of Art.

McQuillen, J. T. (2013). The Uses of Provenance Evidence: Reconstructing the Fifteenth-Century Library of Scheyern Abbey, Bavaria. *The Journal of the Early Book Society for the Study of Manuscripts and Printing History, 16,* 111-148.

Menna, G., & de Vos, M. (2014). Digital Medieval Manuscripts Maps (DMMmaps). Retrieved from http://digitizedmedievalmanuscripts.org/

Metropolitan Museum of Art. (2015). The Collection Online. Retrieved from http://www.metmuseum.org/collection/the-collection-online/search?ft=manuscripts&ao=on&noqs=true

Morgan Library and Museum. (n.d.). Explore the Collection. Retrieved from http://www.themorgan.org/collection

Morville, P. & Rosenfeld, L. (2006). *Information Architecture for the World Wide Web* (3rd ed.). Sebastopol, CA: O'Reilly.

Museum of Modern Art. (2015). Timeline and Map. Retrieved from http://www.moma.org/learn/intnlprograms/timeline

National Library of Ireland. (2015). About the Library. Retrieved from http://www.nli.ie/en/about-the-library.aspx

National Library of Ireland. (n.d.). Catalogue. Retrieved from http://catalogue.nli.ie

National Library of Scotland. (2012). Digital gallery. Retrieved from http://digital.nls.uk/gallery.cfm

National Library of Scotland. (2015). About us. Retrieved from http://www.nls.uk/about-us

National Library of Wales. (2015). View Manuscripts. Retrieved from https://www.llgc.org.uk/collections/digital-gallery/digitalmirror-manuscripts/

Neiswender, C., & Montgomery, E. (2009). Metadata Interoperability – What Is It, and Why Is It Important? *The MMI Guides: Navigating the World of Marine Metadata*. Retrieved from https://marinemetadata.org/guides/mdataintro/mdatainteroperability

New York Public Library. (2015). Digital Collections. Retrieved from http://digitalcollections.nypl.org

Newberry. (n.d.a). Core Collections. Retrieved from https://www.newberry.org/core-collections

Newberry. (n.d.b). Digital Resources and Publications. Retrieved from https://www.newberry.org/digital-resources-and-publications

NISO. (2004). *Understanding Metadata*. Bethesda, MD: National Information Standards Organization Press. Retrieved from http://www.niso.org/publications/press/UnderstandingMetadata.pdf

Niu, J. (2013). Provenance: Crossing boundaries. *Archives and Manuscripts, 41*(2), 105-11. doi: 10.1080/01576895.2013.811426

Norman, J. (2015). HistoryofInformation.com Maps. Retrieved from http://www.historyofinformation.com/maps-simple/?category=Bookbinding&era=&region=

Northwestern University Knight Lab. (2015). StoryMap JS. Retrieved from https://storymap.knightlab.com/

Northwestern University Library. (2012). Schulze-Greenleaf Library. Retrieved from http://digital.library.northwestern.edu/schulze.html

Nylen, J. (Ed.). (2016, August 8). Gallery. Retrieved from https://github.com/d3/d3/wiki/Gallery

OAI. (2002). XML schema to hold provenance information in the "about" part of a record. Retrieved from http://www.openarchives.org/OAI/2.0/guidelines-provenance.htm

OAI. (n.d.). Interoperability through Metadata Exchanged. Retrieved from https://www.openarchives.org/pmh/

OCLC. (2013). Best Practices for CONTENTdm and OAI-PMH compliant repositories: creating sharable metadata. *OCLC Digital Collection Services.* Retrieved from http://www.oclc.org/content/dam/support/wcdigitalcollectiongateway/MetadataBestPractices.pdf

Office for Descriptive Cataloguing Policy Processing Services. (1981). *Bibliographic Description of Rare Books: rules formulated under AACR 2 and ISBD(A) for the descriptive cataloging of rare books and other special materials*. Washington: Library of Congress.

Oldfield, P. (1994). *Ownership of books: an investigation into provenance: an exhibition in the Thomas Fisher Rare Book Library, University of Toronto, 18 July-28 October, 1994*: [Toronto : The Library], 1994.

Oliver, C. (2010). *Introducing RDA: A Guide to the Basics*. Chicago: American Library Association.

Omeka. (2015). Retrieved from http://omeka.org/

Open Knowledge Foundation Labs. (n.d.). TimeMapper. Retrieved from http://timemapper.okfnlabs.org/

Overmier, J. A., & Doak, E. M. (1996). Provenance Records in Rare Book and Special Collections. *Rare Books & Manuscripts Librarianship, 11*(2), 91-99.

Overview of Rare Book Cataloging [Powerpoint slides]. (n.d.). Columbia University. Retrieved from http://www.columbia.edu/cu/libraries/inside/unites/bibcontrol/osmc/rare2.pdf

Pages from the Past. (n.d.). Main Page. Retrieved from http://www.pagesfromthepast.org/

Park, J. & Maszaros, S. (2009). Metadata Object Description Schema (MODS) in Digital Repositories: An Exploratory Study of Metadata Use and Quality *Knowledge Organization, 36*(1), 46-59.

Parker Library on the Web. (2014). Home. Retrieved from https://parker.stanford.edu/parker/actions/page.do?forward=home

Parker Library on the Web. (n.d.a). About Matthew Parker & The Parker Library. Retrieved from https://parker.stanford.edu/parker/actions/page?forward=about_parker

Parker Library on the Web. (n.d.b). About the Project. Retrieved from https://parker.stanford.edu/parker/actions/page?forward=about_project

Pearson, D. (1994). *Provenance Research in Book History: A Handbook*. London: British Library.

Pearson, D. (1997). Exploring and recording provenance: Initiatives and possibilities. *Papers of the Bibliographical Society of America, 91*(4), 505 – 515.

Pearson, D. (2005). Provenance and rare book cataloguing: its importance and its

challenges. In D. J. Shaw (Ed.). *Books and their owners: provenance information and the European cultural heritage : papers presented on 12 November 2004 at the CERL conference hosted by the National Library of Scotland, Edinburgh*. (1-10). London : CERL, 2005.

Peckham, R. D. (n.d.). Consulting Medieval Manuscripts Online. Retrieved from http://www.utm.edu/staff/bobp/vlibrary/mdmss.shtml

Penn Libraries Kislak Center. (n.d.). The Schoenberg Institute for Manuscripts Studies. Retrieved from http://www.library.upenn.edu/kislak/sims/

PennState University Library. (2013). The English Emblem Book Project: Digitized Emblem Books. Retrieved from http://www.libraries.psu.edu/psul/digital/emblem/books.html

PennState University Library. (2014). Digital Collections, The English Emblem Book Project: Project Overview. Retrieved from http://www.libraries.psu.edu/psul/digital/emblem/overview.htm

Pickard, A. J. (2013). *Research Methods in Information, Second Edition*. Chicago: Neal-Schuman.

Polanka, S. (Ed.). (2012). *No Shelf Required 2: Use and Management of Electronic Books*. Chicago: American Library Association.

Posner, E. (1940). Some Aspects of Archival Development since the French Revolution. *The American Archivist, 3*(3), 159-172.

Posner, E. (1972). *Archives in the ancient world*. Cambridge: Harvard University Press.

Powell, R. R., & Connaway, L. S. (2004). *Basic Research Methods for Librarians, Fourth Edition*. Westport, Connecticut: Libraries Unlimited.

Provenir. (n.d.). In *Larousse*. Retrieved from http://www.larousse.fr/dictionnaires/francais/provenir/64641

Rabinowitz, N. (2016). Timemap.js. Retrieved from http://nickrabinowitz.com/portfolio/timemap-js/

RDA and OCLC. (n.d.). OCLC: The world's libraries. Connected. Retrieved from https://oclc.org/en-CA/rda.html

Reimer, S. (2015). Some Useful Websites in Medieval Literature and Manuscript Studies. Retrieved from http://www.ualberta.ca/~sreimer/other/medlinks.htm

Reitz, J. M. (2014a). "digital library". *Online Dictionary for Library and Information Science*. Retrieved from http://www.abc-clio.com/ODLIS/odlis_d.aspx#digitallibrary

Reitz, J. M. (2014b). "provenance". *Online Dictionary for Library and Information Science*. Retrieved from http://www.abc-clio.com/ODLIS/odlis_p.aspx#provenance

Reitz, J. M. (2014c). "Rare Books and Manuscripts Section (RBMS)". *Online Dictionary for Library and Information Science*. Retrieved from http://www.abc-clio.com/ODLIS/odlis_r.aspx#rbms

Reitz, J. M. (2014d). "special collections". *Online Dictionary for Library and Information Science*. Retrieved from http://www.abc-clio.com/ODLIS/odlis_s.aspx#specialcollections

Richard. (2013, October 3). WorldCat Identities Network for Research [Weblog post]. *Toronto Reference Library Blog*. Retrieved from http://torontopubliclibrary.typepad.com/trl/2013/10/worldcat-identies.html

Rimmer, J., Warwick, C., Blandford, A., Gow, J., & Buchanan, G. (2008). An

examination of the physical and the digital qualities of humanities research. *Information Processing and Management 44*, 1374 - 1392.

Roehampton. (2014). Roehampton in the 19th Century. Retrieved from http://209.141.40.101/roehampton-in-the-19th-century/

Rouse, M. (2011, March). MIT License (X11 license or MIT X license). Retrieved from http://whatis.techtarget.com/definition/MIT-License-X11-license-or-MIT-X-license

Rueda-Ramirez, P. J., & Ruiz, M. (2009, August 23-27). *Towards a Provenance Database in the Historic Collection of the Library of the University of Barcelona*. Poster presented at the World Library & Information Congress 75th IFLA General Conference & Council, Milan, Italy. Retrieved from http://www.researchgate.net/publication/46776323_Towards_a_Provenance_Database_in_the_Historic_Collection_of_the_Library_of_the_University_of_Barcelona

Rylands Collections. (n.d.). Browse Rylands Collections. Retrieved from http://luna.manchester.ac.uk/luna/servlet/Manchester~91~1

Sassen, S. (2008). Why Cities Matter. Retrieved from http://www.saskiasassen.com/PDFs/publications/Why-Cities-Matter.pdf.

Scheibe, M. (2010). The 'Biography of Copies': Provenance Description in Online Catalogues. In M. Reed, & B. Wagner (Eds.), *Early Printed Books As Material Objects* (269-278). The Hague, The Netherlands: IFLA Publications.

Schoenberg Center for Electronic Text & Image. (1998). Lawrence J. Schoenberg Collection. Retrieved from http://sceti.library.upenn.edu/ljs/index.cfm

Schoenberg Institute for Manuscript Studies. (2015). Schoenberg Digital Resources. Retrieved from http://schoenberginstitute.org/schoenberg-digital-resources/

Scholar's Lab. (n.d.a). Omeka Plugins. Retrieved from http://scholarslab.org/research/omeka-plugins/

Scholar's Lab. (n.d.b). Neatline: Plot your course in space & time. Retrieved from http://neatline.org/

Schøyen Collection. (n.d.). The Schøyen Collection: manuscripts from around the world spanning 5000 years of human culture & civilisation. Retrieved from http://www.schoyencollection.com/

Selwyn, D. G. (1997). 'Books with manuscript': the case of Thomas Cranmer's library. *The British Library Journal*, *23*(1), 47-59. Retrieved from http://www.jstor.org/stable/42554443

Shaw, D. (2012). Crowdsourcing the Origins of a Medieval Manuscript. Retrieved from https://storify.com/Museocat/crowdsourcing-the-origins-of-a-medieval-manuscript

Shiri, A. (2007). The Use of Metadata in Visual Interfaces to Digital Libraries. In L. Kovács, N. Fuhr, & C. Meghini (Eds.), *Research and Advanced Technology for Digital Libraries,* (489-494). Berlin: Springer Berlin Heidelberg.

Shiri, A. (2008). Metadata-enhanced visual interfaces to digital libraries. *Journal of Information Science, 34*(6), 763-775. doi: 10.1177/0165551507087711

Shiri, A. (2009). Chapter XXVII: Metadata and Metaphors in Visual Interfaces to Digital Libraries. In Y.-L. Theng, S. Foo, D. Goh, & J.-C. Na (Eds.), *Handbook of Research on Digital Libraries: Design, Development, and Impact* (270-277). Hershey, PA: Information Science Reference. doi: 10.4018/978-1-59904-879-6.ch027

Shiri, A. (2011, January 26). *Exploratory Interaction with Digital Libraries Using Visual Thesauri*. Paper presented at the School of Library and Information Studies, University of Alberta, Edmonton, Alberta.

Shiri, A., & Molberg, K. (2005). Interfaces to Knowledge Organization Systems in Canadian Digital Library Collections. *Online Information Review, 29*(6), 604-620.

Smithsonian Libraries. (2004). Incunabula: From the Dibner Library of the History of Science and Technology. Retrieved from http://www.sil.si.edu/digitalcollections/incunabula/

SMU Bridwell Library (2013). Western Manuscripts at Bridwell Library to 1650. Retrieved from https://sites.smu.edu/bridwell/specialcollections/bridwellwesternms/brmscontents.htm

SMU Bridwell Library. (n.d.a). Digital Collection: About. Retrieved from http://digitalcollections.smu.edu/all/bridwell/brinc/index.asp

SMU Bridwell Library. (n.d.b). Fifteenth-Century Printed Books at Bridwell Library. Retrieved from http://digitalcollections.smu.edu/cdm/search/collection/brinc

Society of American Archivists. (2015a). Archives, Personal Papers, and Manuscripts. Retrieved from http://www2.archivists.org/glossary/terms/a/archives-personal-papers-and-manuscripts

Society of American Archivists. (2015b). Describing Archives: A Content Standard, Second Edition. Retrieved from http://files.archivists.org/pubs/DACS2E-2013_v0315.pdf

Society of American Archivists. (2015c). Rules for Archival Description. Retrieved from http://www2.archivists.org/glossary/terms/r/rules-for-archival-description

Standards Committee of the Rare Book and Manuscripts Section (ACRL/ALA). (1988). *Provenance Evidence: Thesaurus for Use in Rare Book and Special Collections Cataloguing*. Chicago: Association of College and Research Libraries.

Stanford University. (2013). Mapping the Republic of Letters: About the Project. Retrieved from http://republicofletters.stanford.edu/index.html

Stein, L. A. (n.d.). Provenance Research at the Archives of American Art. *Smithsonian Institution*. Retrieved from http://www.aaa.si.edu/collections/projects/provenance

Summit, J. (2008). *Memory's Library: Medieval books in early modern England*. Chicago: University of Chicago Press.

Sun, Z., Yue, P, Hu, L., Gong, J., Zhang, L., & Lu, X. (2013). GeoPWProv: Interleaving Map and Faceted Metadata for Provenance Visualization and Navigation. *IEEE Transactions on Geoscience & Remote Sensing, 51*(11), 5135-5136.

Sweeney, S. (2008). The Ambiguous Origins of the Archival Principle of "Provenance". *Libraries & the Cultural Record, 43*(2), 193-213.

Tammaro, A. M. (2016). Heritage Curation in the Digital Age: Professional Challenges and Opportunities. *International Information & Library Review, 48*(2), 122-128. Doi: 10.1080/10572317.2016.1176454

Tayler, A.G. (2006). *Introduction to Cataloging and Classification* (10th ed.). Westport, CT: Libraries Unlimited.

Taylor, D. (2013). *Introduction to RDA* [Powerpoint Slides]. MacEwan University, Edmonton, Alberta.

TEI. (2014). <provenance>. Retrieved from http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-provenance.html

The European Library. (n.d.). Manuscripts and Princes in Medieval and Renaisance Europe. Retrieved from http://www.theeuropeanlibrary.org/tel4/virtual/regia?view=provenance

The Regents of the University of California. (2006). WebGenDB documentation. Retrieved from http://www.lib.berkeley.edu/digicoll/bestpractices/webgendb_documentation.html#users

Theodoridou, M., Tzitzikas, Y., Doerr, M., Marketakis, Y., & Melessanakis, V. (2010). Modeling and querying provenance by extending CIDOC CRM. *Distributed and Parallel Databases, 27*(2), 169-210.

Toronto Public Library. (2016). Toronto in Literature: Neighbourhood Book Lists. Retrieved from http://www.torontopubliclibrary.ca/books-video-music/books/booklists/neighbourhood-list.jsp?utm_source=All&utm_campaign=Issue+%2392+-+Wednesday%2C+February+18%2C+2015&utm_medium=email&utm_source=All&utm_campaign=Issue+%2392+-+Friday%2C+February+20%2C+2015&utm_medium=email

Trinity College Cambridge. (n.d.a). The Digital Library. Retrieved from http://www.trin.cam.ac.uk/library/catalogues/digital-library

Trinity College Cambridge. (n.d.b). Wren Library. Retrieved from http://www.trin.cam.ac.uk/library/wren

Trinity College Dublin. (2011a). Early Printed Books and Special Collections. Retrieved from https://www.tcd.ie/Library/epb/

Trinity College Dublin. (2011b). Manuscripts and Archives Research Library. Retrieved from https://www.tcd.ie/Library/manuscripts/

Trinity College Dublin. (2015). Digital Literary Atlas of Ireland, 1922-1949. Retrieved from http://www.tcd.ie/trinitylongroomhub/digital-atlas/

Trinity College Dublin. (n.d.). Digital Collections, The Library of Trinity College Dublin. Retrieved from http://digitalcollections.tcd.ie/home/

U.S. National Library of Medicine. (n.d.a). About the National Library of Medicine. Retrieved from https://www.nlm.nih.gov/about/

U.S. National Library of Medicine. (n.d.b). Digital Collections. Retrieved from http://collections.nlm.nih.gov/

UBC Library Digital Collections. (n.d.). Retrieved from http://digitalcollections.library.ubc.ca/index.php

UBC Library. (2013). About Rare Books and Special Collections. Retrieved from http://rbsc.library.ubc.ca/about/

UCL Library Services Digital Collections. (2015). Collections. Retrieved from http://digitool-b.lib.ucl.ac.uk:8881/R#special

UCL Library Services. (2014). About us. Retrieved from http://www.ucl.ac.uk/library/special-collections/about-us

University of Aberdeen. (2015). Library, Special Collections and Museums: Collections. Retrieved from http://www.abdn.ac.uk/library/about/special/collections/

University of California, Los Angeles. (n.d.). Hypercities: UCLA. Retrieved from http://hypercities.ucla.edu/

University of Cambridge Museums. (2015). From the early world to out of this world. Retrieved from http://www.fitzmuseum.cam.ac.uk/documents/ucm_leaflet_2015.pdf

University of Chicago Library. (n.d.). Special Collections Research Center, Early Manuscripts. Retrieved from http://www.lib.uchicago.edu/e/scrc/collections/earlymss/

University of Edinburgh. (2015). Special Collections. Retrieved from http://www.ed.ac.uk/schools-departments/information-services/library-museum-gallery/crc/collections/special-collections

University of Glasgow. (n.d.a). Special collections: About us. Retrieved from http://www.gla.ac.uk/services/specialcollections/aboutus/

University of Glasgow. (n.d.b). Special collections: Digitised books, manuscripts and photographs. Retrieved from http://www.gla.ac.uk/services/specialcollections/digitisedbooksmss/

University of Iowa Libraries. (2015). Medieval Manuscripts Digital Collection. Retrieved from http://digital.lib.uiowa.edu/mmc/

University of Leeds. (2015). Special Collections, The University Library. Retrieved from http://library.leeds.ac.uk/special-collections

University of Leeds. (n.d.a). About this resource. Retrieved from http://library.leeds.ac.uk/special-collections/view/831/about_this_resource

University of Leeds. (n.d.b). Medieval Illuminated Manuscripts Digital Resource. Retrieved from http://library.leeds.ac.uk/special-collections/view/830/medieval_manuscripts_home

University of Madison-Wisconsin Digital Collections. (2011). University of Wisconsin-Madison Special Collections. Retrieved from http://uwdc.library.wisc.edu/collections/UWSpecColl

University of Manitoba Libraries. (2015a). Dysart Memorial Collection Virtual Exhibition. Retrieved from http://umanitoba.ca/libraries/units/archives/collections/rarebooks/dysart/exhibition.html

University of Manitoba Libraries. (2015b). Mission and Mandate. Retrieved from http://umanitoba.ca/libraries/units/archives/mandate.html

University of North Carolina. (n.d.a). About the Collection. Retrieved from http://dc.lib.unc.edu/cdm/about/collection/mackinney/

University of North Carolina. (n.d.b). About Loren C. MacKinney. Retrieved from http://dc.lib.unc.edu/cdm/aboutmackinney/collection/mackinney/

University of North Carolina. (n.d.c). The MacKinney Collection of Medieval Medical Illustrations. Retrieved from http://dc.lib.unc.edu/cdm/search/collection/mackinney

University of Pennsylvania. (n.d.a). Penn in Hand. Retrieved from http://dla.library.upenn.edu/dla/medren/index.html

University of Pennsylvania. (n.d.b). Penn in Hand: Overview. Retrieved from http://dla.library.upenn.edu/dla/medren/ancillary.html?id=collections/rbm/mss/mss

University of Saskatchewan. (2005). Remaking the Book: Digitally Reconstructing the Otto Ege Manuscript Portfolios. Retrieved from http://library2.usask.ca/ege/

University of Saskatchewan. (n.d.). University Archives & Special Collections. Retrieved from http://library.usask.ca/archives/index.php

University of St Andrews Library. (n.d.). Digital Collections. Retrieved from https://arts.st-andrews.ac.uk/digitalhumanities/node/2

University of St Andrews. (2015). History of the Special Collections Division. Retrieved from http://www.st-andrews.ac.uk/library/specialcollections/aboutus/historyofspecialcollections/

University of Sydney. (2016). Timemap: Time-based interactive mapping. Retrieved from http://sydney.edu.au/arts/timemap/index.shtml

University of Toronto Thomas Fisher Rare Book Library. (2014a). Retrieved from http://fisher.library.utoronto.ca/

University of Toronto Thomas Fisher Rare Book Library. (2014b). Fisher Digitized Manuscripts. Retrieved from http://fisher.library.utoronto.ca/resources/digitized-manuscripts

University of Toronto. (n.d.). British Armorial Bindings. Retrieved from https://armorial.library.utoronto.ca/

University of Victoria Libraries. (2015). Medieval and early modern manuscripts collections. Retrieved from http://www.uvic.ca/library/locations/home/spcoll/collections/medieval/inventory.php#manuscripts

University of Virginia Library Scholar's Lab. (n.d.). The Battle of Chancellorsville. Retrieved from http://hotchkiss.scholarslab.org/neatline-exhibits/show/battle-of-chancellorsville/fullscreen

University of Washington Libraries. (n.d.a). Digital Collections, Browsing items in: Medieval and Historical Manuscripts. Retrieved from http://digitalcollections.lib.washington.edu/cdm/search/collection/mhm

University of Washington Libraries. (n.d.b). Digital Collections: Medieval and Historical Manuscripts Collection. Retrieved from http://content.lib.washington.edu/mhmweb/index.html

van Heel, J. (2011). Some Notes on Research into the Provenance of Medieval Books. *Quaerendo, 41*, 256-262.

Van Verchum, M., & Pugin, L. (2014, July 12-18). *Visualising large data sets of music bibliographic records*. Paper presented at the annual International Association of Music Libraries Conference, Antwerp, Belgium.

Villanova Library. (n.d.). VuFind. Retrieved from http://vufind-org.github.io/vufind/

Vullo, G. (2010). Squaring the Circle: A Comparative Perspective on Digital Library Evaluation. In A. Katsirikou, & C. Skiadas (Eds.), *New Trends in Qualitative and Quantitative Methods in Libraries* (353-358). New Jersey: World Scientific.

W3C. (2013). PROV-Overview: An Overview of the PROV Family of Documents. Retrieved from http://www.w3.org/TR/prov-overview/

W3C. (2015). Linked Data. Retrieved from http://www.w3.org/standards/semanticweb/data

Walters Art Museum. (n.d.). The Digital Walters: Digitized Walters Manuscripts. Retrieved from http://www.thedigitalwalters.org/01_ACCESS_WALTERS_MANUSCRIPTS.html

Warwick Digital Collections. (n.d.). Library Special Collections. Retrieved from http://contentdm.warwick.ac.uk/cdm/landingpage/collection/spc

Warwick The Library. (2015). Rare books & special collections. Retrieved from http://www2.warwick.ac.uk/services/library/using/rare-books/

Weimer, L. (2007). Pathways to Provenance: DACS and Creator Descriptions. *Journal of Archival Organization, 5*(1/2), 33-48. doi: 10.1300/J201v05n01_03

Wellcome Trust. (2015a). Digital collection: Medieval and Early Modern manuscripts. Retrieved from http://wellcomelibrary.org/collections/browse/collections/digwms

Wellcome Trust. (2015b). What We Do. Retrieved from http://wellcomelibrary.org/what-we-do/

Western Medieval Manuscripts. (n.d.). University of Edinburgh Collections. Retrieved from http://images.is.ed.ac.uk/luna/servlet/UoEwmm~1~1

Western Michigan University Libraries. (n.d.). Medieval Document Collection. Retrieved from http://luna.library.wmich.edu:8180/luna/servlet/WMUwmu~77~77

Whittaker, B., & Thomas, L. (2009). *Special Collections 2.0: new technologies for rare books, manuscripts, and archival collections*. Santa Barbara, CA: Libraries Unlimited.

Wickens, G. E. (2004). *Economic Botany: Principles and Practices*. Dordrecht, Netherlands: Kluwer Academic Publishers.

Wiggins, B. (2012, March 2). Library of Congress Announces Its Long-Range RDA Training Plan. *Library of Congress Cataloging and Acquisitions*. Retrieved from http://www.loc.gov/catdir/cpso/news_rda_implementation_date.html

Wildemuth, B. M. (Ed.). (2009). *Applications of Social Research Methods to Questions in Information and Library Science*. Westport, Connecticut: Libraries Unlimited.

Williford, C., & Henry, C. (2012a). Case Studies: Digging into the Enlightenment: Mapping the Republic of Letters. In *One Culture. Computationally Intensive Research in the Humanities and Social Sciences: A Report on the Experiences of*

First Respondents to the Digging Into Data Challenge. Pub151: Council on Library and Information Resources. Retrieved from https://www.clir.org/pubs/reports/pub151/case-studies/enlightenment

Williford, C., & Henry, C. (2012b). Reflecting and Looking Ahead. In *One Culture. Computationally Intensive Research in the Humanities and Social Sciences: A Report on the Experiences of First Respondents to the Digging Into Data Challenge*. Pub151: Council on Library and Information Resources. Retrieved from https://www.clir.org/pubs/reports/pub151/3.-reflecting-and-looking-ahead

Woodley, M. S. (2008). Crosswalks, Metadata Harvesting, Federated Searching, Metasearching: Using Metadata to Connect Users and Information. *Introduction to Metadata*. Retrieved from http://getty.edu/research/publications/electronic_ publications/intrometadata/path.pdf

Yadagiri, N., & Ramesh, P. (2013). *Semantic Web and the Libraries: An Overview*. International Journal of Library Science, 7(1), 80-94. Retrieved from http://www.semantic-web-journal.net/sites/default/files/swj321.pdf

Yale University Beinecke Rare Books & Manuscript Library. (2013a). Beinecke Rare Book & Manuscript Library. Retrieved from http://beinecke.library.yale.edu/

Yale University Beinecke Rare Books & Manuscript Library. (2013b). Mission. Retrieved from http://beinecke.library.yale.edu/about/mission

Zaino, J. (2013, February 13). The Future of Libraries, Linked Data and Schema.Org Extensions [web log post]. Retrieved from http://www.dataversity.net/the-future-of-libraries-linked-data-and-schema-org-extensions/

Zick, G. (2009). Digital Collections: History and Perspectives. *Journal of Library Administration, 49*, 687-693.

Appendix A: Provenance Metadata in Major Standards and Schemas

At the outset of this thesis, one of the formative research questions was to identify the standards that exist in provenance metadata, how they differ, and to examine some reasons that multiple standards might exist. Examining how provenance is captured in metadata standards and schemas provides insight into the information available to end-users, the depth and complexity of that information, and challenges libraries might be facing in providing provenance information. This appendix provides an overview of the interpretation of provenance within major standards and schemas.

The examination of provenance within metadata standards and schemas includes identifying the overall function of the standard, the primary knowledge institution associated with the standard, the name of the element itself used to identify provenance, and the scope and definition of that element.

Numerous standards and schemas exist; however, only some of the most important or relevant ones are addressed here, in *Table A1*. Standards and schemas have been organized according to Elings & Waibel's classification, and include older standards and schemas that were influential enough to still be important today. More recent interpretations of standards and schemas have also been included.

Table A1

Summary of Provenance Metadata in Major Standards and Schemas

Metadata Standard or Schema	Primary Knowledge Institution(s)	Provenance Metadata Element(s)	Scope and Definition
Data Structure			
Categories for the Description of Works of Art (CDWA)	Museums	Ownership/ Collection History	- Includes all owners who have possessed the item in question, including how and when transfer of ownership occurred. If sections of history are missing, this should be noted. - Includes other contextual information such as item cost, legal status, current owner, dates of ownership, identifying numbers, citations for consulted resources. Information can be in a prose paragraph, or parceled into sub-elements. - Information on identifying marks, and other provenance evidence are captured in "Materials and Techniques" and "Inscription/Marks" (Harpring, 2009)
MAchine-Readable Cataloguing (MARC)	Libraries	Field 3xx; Fields 541 and 561; Field 590; Fields 700 and 710; Field 852.	- Extent of an item; - Note fields for recent acquisition and provenance history, including owner names and types of evidence; - Note field for copy-specific information; - Personal name fields with relator terms ($e) and relator codes ($4); - Copy-specific information of individual copies of a book (Lundy, 2008).
Metadata Object Description Schema (MODS)	Libraries	<note type= "ownership">	- Includes information relevant to provenance, equivalent to MARC Field 561 (Library of Congress, 2010).

Special Collections and Archives Code (SPAC)	Libraries	MARC field 590 or 901 (note fields)	- Suggested Metadata item that could be added to any collection to identify what collection an item may have come from (Buchanan, 2011).
Encoded Archival Description (EAD)	Archives	<custodhist> Provenance; <acqinfo> Acquisition Information	- Provenance is "information about the chain of ownership of the materials being described, before they reached the immediate sources of acquisition. Both physical possession and intellectual ownership can be described, providing details of change of ownership and/or custody that may be significant in terms of authority, integrity, and interpretation." (DCMI, 2002, EAD). - Acquisition information is "the immediate source of the materials being described, and the circumstances under which they were received." (DCMI, 2002, EAD).
Data Content			
Cataloguing Cultural Objects (CCO)	Museums	N/A	- Cataloguers must refer and use CDWA for provenance information (Harpring, 2009).
International Standard Bibliographic Description (Antiquarian) (ISBD(A))	Libraries	Note field	- Provenance information should be captured in a narrative or prose-like fashion, according to guideline 7.9. Broadly covers history of ownership (Lundy, 2008; Office for Descriptive Cataloguing Policy Processing Services, 1981).
Anglo-American Cataloguing Rules, Second Edition (AACR2)	Libraries	Note fields	- Rules 1.7B20 for "Copy being described", 2.7B20 for "Descriptive details of particular copy described", and 2.18F1 for "Binding information" suggest provenance information be captured in note fields (AACR2, 2002; Lundy, 2009).

			- Recommended use of additional guidelines such as Descriptive Cataloguing of Rare Materials (Books) (DCRM(B)), *Guidelines for the Cataloguing of Rare Books*, and the *Provenance Evidence Thesaurus* to supplement AACR2, and use controlled vocabulary (Overview of Rare Book Cataloguing, n.d.).
Descriptive Cataloguing of Rare Materials (Books) (DCRM(B))	Libraries	Note fields	- DCRM(B) should be used to supplement AACR2 (it cannot be used separately from AACR2) - Rule 7A1.4 says "Notes may also be made to justify added entries intended for special fields of personal or corporate names, titles, genres/forms, physical characteristics, provenance, etc. Whenever possible use terms taken from lists of controlled vocabularies when making such notes and added entries." (DCRB, 1991; Library of Congress, 2007). - Advised to use *Provenance Evidence Thesaurus* for vocabulary (DCRB, 1991; Library of Congress, 2007). - Rule 7B19.2 Suggests that provenance should be captured according to *Guidelines for the Cataloguing of Rare Books* (From DCRB to DCRM(B), n.d.). - Other contextual provenance evidence is explained through other rules not explicitly linked to provenance (From DCRB to DCRM(B), n.d.).
Guidelines for the Cataloguing of Rare Books	Libraries	N/A	- *Guidelines* is meant to be a supplement to AACR2 and DCRM(B). - *Guidelines* does not provide new fields for describing provenance, but rather provides instructions for how best to fill current fields, such as MARC 561 (through free-text), 700/710 (using name authorities for owners), and 655 (using

			headings for provenance evidence) (Hillyard & Pearson, 2007). - *Guidelines* also encourages the use of thesauri and name authorities (Hillyard & Pearson, 2007). - Multi-level approaches are encouraged for libraries depending on their expertise or ability to do rare book cataloguing, with the minimum level suggesting capturing former ownership with associated dates, and higher levels including more detailed descriptions of provenance evidence (Hillyard & Pearson, 2007).
Resources Description and Access (RDA)	Libraries	Notes; Custodial History of Item	- Unlike its predecessor, AACR2, RDA provides a specific rule for provenance (2.17), which broadly defines custodial history as "A record of previous owernship or custodianship of an item" (Chartered Institute of Library and Information Professionals (Great Britain), Joint Steering Committee for Development of RDA, American Library Association, & Canadian Library Association, 2013, Rule 2.17). - Provenance information can also be captured in notes. However, increased emphasis is placed on relationships in RDA (Oliver, 2012), suggesting that semantic-web like interconnections can be made through RDA cataloguing.
Dublin Core	Libraries	N/A; Description	- Dublin Core does not currently cover provenance through a specific element, although such information is often included in the element "Description". - The term dc:provenance has been suggested, as "A statement of any changes in the ownership and custody of the resource that are significant for its authenticity, integrity, or interpretation." (DCMI, 2002, "provenance").

General International Standard for Archival Description, Second Edition (ISAD(G))	Archives	Archival History; Sources of Acquisition; Related Units	- Provenance is defined as "the relationship between records and the organizations or individuals that created, accumulated, and/or maintained and used them in the conduct of personal or corporate activity" (International Council on Archives, 2000, p.11). - Emphasis on relationships of owners to materials, instead of a simple list of owners. However, no specification of other contextual information (e.g. dates) is made. - Rule 3.2.3 "Archival History" broadly includes information on the history of a collection. - Rule 3.2.4 addresses the immediate source of acquisition of a collection. - Rule 3.5.3 captures information on related materials to the objects being captured, including relation through provenance (International Council on Archives, 2000). - ISAD(G) standardizes *finding aids*, as opposed to individual records for material.
Archives, Personal Papers, and Manuscripts (APPM)	Archives	Provenance; Notes	- APPM has been mostly replaced by DACS. It is used to capture metadata on a collection-level (Society of American Archivists, 2015a). - APPM is based upon AACR2, and notes that provenance is not adequately covered in the latter. Consequently, added entries on biographical or historical sketches on the creator of the materials, scope entries, and custodial history/sources of acquisition are allowed, as well as heavy use of notes (Hensen, 1983).
Describing Archives: A Content Standard (DACS)	Archives	Name of Creators; Collection Description;	- According to Rule 2.6, creators must be described in DACS, as well as materials (Society of American Archivists, 2015b).

		Administrative/ Biographical History; Related Materials; Date Notes	- This can also be elaborated upon in Rule 2.7 Administrative/Biographical History, which is optional. - Materials related to a description through provenance must also be highlighted, according to Rule 6.3 (Society of American Archivists, 2015b). - Rule 2.4 also states that dates are required in the DACS standard, related to both the collection, and its provenance (Society of American Archivists, 2015b). - In Rule 7.1.8, Notes elements can also be used for reconstructing provenance (Society of American Archivists, 2015b). - In DACS, provenance is also a founding principle for organizing a collection (Society of American Archivists, 2015b).
Rules for Archival Description (RAD)	Archives	Scope; Administrative History/Biographical Sketch; Custodial History	- Rule 13.0A1 describes Scope as allowing for provenance information, if relevant to the composition of the collection (e.g. the collection was created through the ownership of X person). - Rule 13.7B1 describes Administrative History as providing background about the previous owners of the collection.Rule 13.7B2 describes the same thing, but related to person(s) or families who owned a collection, instead of organizations. - Rule 13.7C suggests that a chain of ownership should be captured, in chronological order, of a collection, where possible. - RAD is an extenstion of AACR2 (Society of American Archivists, 2015c), developed by Canadian Archivists, used to capture *fonds* instead of individual items.

ContentDM	Digital collections	Provenance	- Provenance information of any format is included in a separate element within systems using ContentDM platforms, known as "Provenance".
Federal Geographic Data Committee (FGDC)'s Geospatial Metadata	Geographic systems	N/A	- Geospatial metadata is not commonly used for metadata, but has an interest in capturing both temporal and geographic data related (most often) to geographic-based data (Liu, 2007). This metadata is often used in software such as Geographic Information Systems (GIS) (Liu, 2007). These developments are of note due to the inherent geographic and temporal information of provenance.
Data Format			
Text Encoding Initiative (TEI)	Humanities, Social Science Disciplines	<provenance>	- The provenance tag can be used to "[contain] any descriptive or other information concerning a single identifiable episode during the history of a manuscript or manuscript part, after its creation but before its acquisition" (TEI, 2014, para.3). - Manuscripts are included, but not other rare materials. - Information is usually in a note-like or narrative format.
Categories for the Description of Works of Art (CDWA)-Lite	Museums	<cdwalite: inscriptions>; <cdwalite: labelRelatedWork>; <cdwalite: locationName type= formerLocation> / <cdwalite: locationName type= formerGeographic>	- Most Ownership/Collecting History administrative metadata elements are not covered in CDWA-Lite (Baca, Clarke, Eklund, Gilliland, Harpring, Woodley, & O'Keefe, 2014) - Tags related to provenance, such as inscriptions on an object, works that might be related through sharing a previous collection, or the name of former collections or institutions / geographic locations of an item can be described using CDWA-Lite (Harpring, 2009).
Metadata Encoding and Transmission Standard (METS)	Libraries, Archives	<digiprovMD>	- In METS, provenance refers to "any preservation-related actions taken on the various files which comprise an digital object […] or, in the case of born digital materials, the files' creation. In short, digital provenance should be used to record

			information that allows both archival/library staff and scholars to understand what modifications have been made to a digital object and/or its constituent parts during its life cycle" (Library of Congress, 2014, Element <digiprovMD>) - Allows for the metadata to be represented according to standards, or as locally developed descriptions. (Library of Congress, 2014). - Rare book provenance is not covered in METS.
Resource Description Framework (RDF)	Semantic Web	PROV	- Provenance relates to the workflow of data, for the most part in RDF (W3C, 2013), as outlined by definitions in computer science (see "Provenance Theory").
Data Exchange			
Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)	Museums, Libraries, Archives	originDescription within an XML <provenance> container	- Allows for capture of the chain of ownership of an item. - OAI-PMH recognizes that provenance cannot always be captured in a simple manner.
CIDOC Conceptual Reference Model (CIDOC-CRM)	Museums	N/A	- Suggestions for an extension allowing for capture of provenance of *digital items* have been made, and their subsequent use in the Semantic Web (Theodoridou, Tzitzikas, Doerr, Marketakis, and Melessanakis, 2010). - Provenance of physical objects has not yet been covered.
VRA Core	Museums (for artwork)	Description	- Any provenance information related to the image being described can be written in a free-text format in the Description field (Library of Congress, n.d.c)
Lightweight Information Describing Objects (LIDO) Schema	Museums	eventType	- A broad category that can be used to capture an event related to the object being described, such as change in ownership (LIDO, 2020) - Element can be repeated multiple times.
Table A1			

Appendix B: Sample Data for Visualization

The second half of the study of this thesis investigates currently available software for visualizing provenance metadata. It seeks to ascertain whether or not visualization is a viable means of contemporary use of metadata. Visualization could provide users with new insights while also providing means of standardizing metadata. In order to test five identified visualization software, test provenance metadata was necessary. *Appendix B* provides the breakdown of ten item records as provenance metadata in *Table B1*, used for visualization case studies in this thesis.

The ten manuscripts in *Table B1* were chosen using purposive sampling based upon the knowledge of the researcher. Ten items provide a small enough number to be able to manage across four to seven tools, but also generates enough data that trends, conclusions, and results can be drawn from testing.

Manuscripts identified and used within the study all came from the digital Lawrence J. Schoenberg Collection from the University of Pennsylvania Libraries, which includes full digital copies of the works in question. This particular collection was chosen not only for its variety of codices, but because emphasis has been placed on provenance in metadata capture. This means the collection came with known provenance, and additional research was not necessary. This was necessary due to time constraints.

Using manuscripts that come from one collection was ideal for experimenting with visualization – it is known that all the items chosen in the sample interact at some point by coming into the same collection. Ability to highlight these relationships is ideal in visualization. All manuscripts chosen were further selected by criteria set out within the study – all items were of Western origin (ensured through the selection of

manuscripts written in western languages), and from roughly the same time period of the fifteenth and sixteenth centuries.

In spite of this, metadata had to be cleaned in order to be useable in visualization testing (Van Verchum & Pugin, 2014) so that formatting was consistent (e.g. that place names or time periods were standard), and the three axes of owners, time, and place could be isolated. The metadata gathered in *Table B1* has therefore been cleaned and format has been made consistent so that the metadata is interoperable and identified facets can be used within visualization.

Table B1

Sample Data Collected for Testing of Provenance as Visualization

Project Assigned Manuscript Identifier	Relevant Manuscript Metadata						
A	*Title*	[L'arte del navegare] [manuscript]					
	Author	--					
	Date of Publication	1464-1465					
	Location of Publication	Venice?, Italy					
	URL	http://hdl.library.upenn.edu/1017/d/medren/4844091					
	Provenance Information by Axis	*Owner*[32] *1*	Frederick North, fifth Earl of Guilford	*Provenance Date(s) 1*	?-13 Dec. 1830	*Provenance Location 1*	Guilford, England
		Owner 2	John Lee	*Provenance Date(s) 2*	18??- 8 Nov. 1888	*Provenance Location 2*	London, England
		Owner 3	Sotheby's	*Provenance Date(s) 3*	10 Jul. 1968	*Provenance Location 3*	London, England
		Owner 4	Sotheby's	*Provenance Date(s) 4*	7 Dec. 2004	*Provenance Location 4*	London, England
		Owner 5	Lawrence J. Schoenberg (and Barbara Brizdle)	*Provenance Date(s) 5*	7 Dec. 2004 - 2011	*Provenance Location 5*	Philadelphia, Pennsylvania

[32] For purposes of simplicity, the term owner has been used here to refer to owners, institutions, corporations, and auction houses.

		Owner 6	Rare Book & Manuscript Library University of Pennsylvania	*Provenance Date(s) 6*	2011 - Present	*Provenance Location 6*	Philadelphia, Pennsylvania
B	*Title*	[Medical and astronomical miscellany] [manuscript]					
	Author	--					
	Date of Publication	circa 1446					
	Location of Publication	Germany					
	URL	http://hdl.library.upenn.edu/1017/d/medren/4838166					
	Provenance Information by Axis	*Owner 1*	Nicolas-Antoine Labbey-de-Billy	*Provenance Date(s) 1*	1753-1825	*Provenance Location 1*	Besançon, France
		Owner 2	Ader, Picard, Tajan	*Provenance Date(s) 2*	1976	*Provenance Location 2*	Paris, France
		Owner 3	Sotheby's	*Provenance Date(s) 3*	1982	*Provenance Location 3*	London, England
		Owner 4	Jörn Günther	*Provenance Date(s) 4*	1997	*Provenance Location 4*	Stalden, Switzerland
		Owner 4b	Jörn Günther (with Ursus Books Ltd.)	*Provenance Date(s) 4b*	1997	*Provenance Location 4b*	New York, United States
		Owner 4c	Jörn Günther (with Roth Horowitz, Ferrini & Biondi)	*Provenance Date(s) 4c*	2002	*Provenance Location 4c*	Los Angeles, United States
		Owner 5	Sam Fogg	*Provenance Date(s) 5*	2003	*Provenance Location 5*	London, England

		Owner 6	Lawrence J. Schoenberg (and Barbara Brizdel)	*Provenance Date(s) 6*	May 2003 - 2011	*Provenance Location 6*	Philadelphia, Pennsylvania
		Owner 7	Rare Book & Manuscript Library University of Pennsylvania	*Provenance Date(s) 7*	2011 - Present	*Provenance Location 7*	Philadelphia, Pennsylvania
C	*Title*	[Alchemical compendium] [manuscript]					
	Author	Hayniger, Georg					
	Date of Publication	circa 1476					
	Location of Publication	Vienna, Austria					
	URL	http://hdl.library.upenn.edu/1017/d/medren/5783938					
	Provenance Information by Axis	*Owner 1*	Johann David Baier	*Provenance Date(s) 1*	late 1600s	*Provenance Location 1*	Nuremberg, Germany
		Owner 1b	Johann David Baier	*Provenance Date(s) 1b*	1700s	*Provenance Location 1b*	Jena, Germany
		Owner 1c	Johann David Baier	*Provenance Date(s) 1c*	1730s and 1740s	*Provenance Location 1c*	Altdorf, Germany
		Owner 2	P. Fraenkel	*Provenance Date(s) 2*	1999	*Provenance Location 2*	Geneva, Switzerland
		Owner 3	Sam Fogg Ltd.	*Provenance Date(s) 3*	2000	*Provenance Location 3*	London, England
		Owner 4	Lawrence J. Schoenberg (and Barbara Brizdle)	*Provenance Date(s) 4*	2000-2012	*Provenance Location 4*	Philadelphia, Pennsylvania

		Owner 5	Rare Book & Manuscript Library University of Pennsylvania	*Provenance Date(s) 5*	2012 - Present	*Provenance Location 4*	Philadelphia, Pennsylvania
D	*Title*	[Choir psalter gatherings] [manuscript]					
	Author	Catholic Church					
	Date of Publication	15--					
	Location of Publication	Spain					
	URL	http://hdl.library.upenn.edu/1017/d/medren/4270055					
	Provenance Information by Axis	*Owner 1*	Sotheby's	*Provenance Date(s) 1*	1978	*Provenance Location 1*	London, England
		Owner 2	Lawrence J. Schoenberg (and Barbara Brizdle)	*Provenance Date(s) 2*	1978 - 2009	*Provenance Location 2*	Philadelphia, Pennsylvania
		Owner 3	Rare Book & Manuscript Library University of Pennsylvania	*Provenance Date(s) 3*	2009 - Present	*Provenance Location 3*	Philadelphia, Pennsylvania
E	*Title*	Abscondita naturae [manuscript]: collecta ab diversis philosoph[orum] et medicorum voluminibus / scriptum in inclita civitate Crem[onensi] per Hieronymum Vidam					
	Author	Vida, Marco Girolamo					
	Date of Publication	2545					
	Location of Publication	Cremona, Italy					
	URL	http://hdl.library.upenn.edu/1017/d/medren/5807343					

	Provenance Information by Axis	*Owner 1*	Ludovico Settala	*Provenance Date(s) 1*	late 1500s – 1633	*Provenance Location 1*	Milan, Italy
		Owner 2	Sebastiano Calvi (gifted by Giacomo Filippo Scaravaggi)	*Provenance Date(s) 2*	1685	*Provenance Location 2*	Milan, Italy
		Owner 3	Martayan Lan	*Provenance Date(s) 3*	? - 2003	*Provenance Location 3*	New York, United States
		Owner 4	Lawrence J. Schoenberg (and Barbara Brizdle)	*Provenance Date(s) 4*	2003 - 2012	*Provenance Location 4*	Philadelphia, Pennsylvania
		Owner 5	Rare Book & Manuscript Library University of Pennsylvania	*Provenance Date(s) 5*	2012 - Present	*Provenance Location 5*	Philadelphia, Pennsylvania
F	*Title*	Herbal … [etc.] [manuscript]					
	Author	--					
	Date of Publication	1520 ; 1600					
	Location of Publication	Italy ; England					
	URL	http://hdl.library.upenn.edu/1017/d/medren/4849768					
	Provenance Information by Axis	*Owner 1*	Lanthan family	*Provenance Date(s) 1*	1501-1600?	*Provenance Location 1*	Lancashire, England
		Owner 2	Thomas and Edward Syddell	*Provenance Date(s) 2*	late 1500s	*Provenance Location 2*	Lancashire, England
		Owner 3	Bonhams	*Provenance Date(s) 3*	Dec. 1995	*Provenance Location 3*	London, England

		Owner 4	Sam Fogg Ltd.	*Provenance Date(s) 4*	June 1996	*Provenance Location 4*	London, England
		Owner 5	Lawrence J. Schoenberg (and Barbara Brizdle)	*Provenance Date(s) 5*	June 1996 - 2012	*Provenance Location 5*	Philadelphia, Pennsylvania
		Owner 6	Rare Book & Manuscript Library University of Pennsylvania	*Provenance Date(s) 6*	2012 - Present	*Provenance Location 6*	Philadelphia, Pennsylvania
G	*Title*	Crispi Sallustii de Catheli[n]e co[n]jurac[i]o[n]e liber [manuscript] ; Jugurtinum bellum ei[us]de[m]. Bellum Catilinae					
	Author	Sallust					
	Date of Publication	between 1455 and 1465?					
	Location of Publication	Padua?, Italy					
	URL	http://hdl.library.upenn.edu/1017/d/medren/4649182					
	Provenance Information by Axis	*Owner 1*	Maffei family	*Provenance Date(s) 1*	1501-1600?	*Provenance Location 1*	Volterra, Italy
		Owner 2	Clemented Carilli	*Provenance Date(s) 2*	1601-1700?	*Provenance Location 2*	Volterra, Italy
		Owner 3	Giuseppe Antonio Sccardini	*Provenance Date(s) 3*	1666	*Provenance Location 3*	Volterra, Italy
		Owner 4	Christie's	*Provenance Date(s) 4*	2000	*Provenance Location 4*	London, England
		Owner 5	Lawrence J. Schoenberg (and Barbara Brizdle)	*Provenance Date(s) 5*	June 2000 - 2010	*Provenance Location 5*	Philadelphia, Pennsylvania

		Owner 6	Rare Book & Manuscript Library University of Pennsylvania	*Provenance Date(s) 6*	2010 - Present	*Provenance Location 6*	Philadelphia, Pennsylvania
H	*Title*	[Scientific miscellany] [manuscript]					
	Author	Fentryer, Imbert					
	Date of Publication	1511					
	Location of Publication	France					
	URL	http://hdl.library.upenn.edu/1017/d/medren/5034562					
	Provenance Information by Axis	*Owner 1*	Christian Friderich	*Provenance Date(s) 1*	1633	*Provenance Location 1*	--
		Owner 2	Anthelme-Michel-Laurent de Migieu, marquis de Savigny	*Provenance Date(s) 2*	1760	*Provenance Location 2*	Savigny, France
		Owner 3	Rodolphe Chamonal, Biennale des Antiquaires	*Provenance Date(s) 3*	1988	*Provenance Location 3*	Paris, France
		Owner 4	John D. Stanitz	*Provenance Date(s) 4*	? - 1997	*Provenance Location 4*	Cleveland, Ohio
		Owner 5	Lawrence J. Schoenberg (and Barbara Brizdle)	*Provenance Date(s) 5*	1997 - 2011	*Provenance Location 5*	Philadelphia, Pennsylvania
		Owner 6	Rare Book & Manuscript Library University of Pennsylvania	*Provenance Date(s) 6*	2011 - Present	*Provenance Location 6*	Philadelphia, Pennsylvania
I	*Title*	Prenosticatio [manuscript]. Prognosticatio.					

	Author	Litchtenberger, Johannes					
	Date of Publication	circa 1500					
	Location of Publication	Nuremberg?, Germany					
	URL	http://hdl.library.upenn.edu/1017/d/medren/5806807					
	Provenance Information by Axis	*Owner 1*	Veit Engelhardt	*Provenance Date(s) 1*	--	*Provenance Location 1*	Nuremberg, Germany
		Owner 2	Jörn Günther (and Bruce Ferrini)	*Provenance Date(s) 2*	1999 - 2000	*Provenance Location 2*	Stalden, Switzerland
		Owner 2b	Jörn Günther (and the John J. Burns Library, Boston College)	*Provenance Date(s) 2b*	2000	*Provenance Location 2b*	Boston, United States
		Owner 3	Sam Fogg Ltd.	*Provenance Date(s) 3*	2000?-Dec. 2002	*Provenance Location 3*	London, England
		Owner 4	Lawrence J. Schoenberg (and Barbara Brizdle)	*Provenance Date(s) 4*	Dec. 2002 - 2012	*Provenance Location 4*	Philadelphia, Pennsylvania
		Owner 5	Rare Book & Manuscript Library University of Pennsylvania	*Provenance Date(s) 5*	2012 - Present	*Provenance Location 5*	Philadelphia, Pennsylvania
J	*Title*	Dos discursos de la cifra [manuscript]					
	Author	--					
	Date of Publication	circa 1600					
	Location of Publication	Spain					

	URL	http://hdl.library.upenn.edu/1017/d/medren/4919302					
	Provenance Information by Axis	*Owner 1*	Martayan Lan	*Provenance Date(s) 1*	2001	*Provenance Location 1*	New York, United STates
		Owner 2	Lawrence J. Schoenberg (and Barbara Brizdle)	*Provenance Date(s) 2*	2001-2011	*Provenance Location 2*	Philadelphia, Pennsylvania
		Owner 3	Rare Book & Manuscript Library University of Pennsylvania	*Provenance Date(s) 3*	2011-Present	*Provenance Location 3*	Philadelphia, Pennsylvania

Note. All titles pulled from the Lawrence J. Schoenberg Collection at the Rare Book & Manuscript Library, University of Pennsylvania.

Table B1

Appendix C: Descriptive and Digital Factor Graphs and Provenance Metadata

This appendix provides graphical representation of the relationship between descriptive and digital factors and provenance metadata, as discussed in "Provenance Metadata in Digital Special Collections".

Descriptive Factors and Provenance Metadata

Aspects such as whether or not a digital collection has visible ties to archival institutions, the age of the institution, the age of the digital collection, the size of the digital collection, the type of institution the digital collection originates from, and the geographic location of the originating institution can all be considered descriptive factors of a digital collection. These aspects inform and describe the collection in some way. The potential correlation of these descriptive factors and the presence of provenance metadata, as well as the axes available within that metadata (see *Table* 5), is represented by graphs within this section.

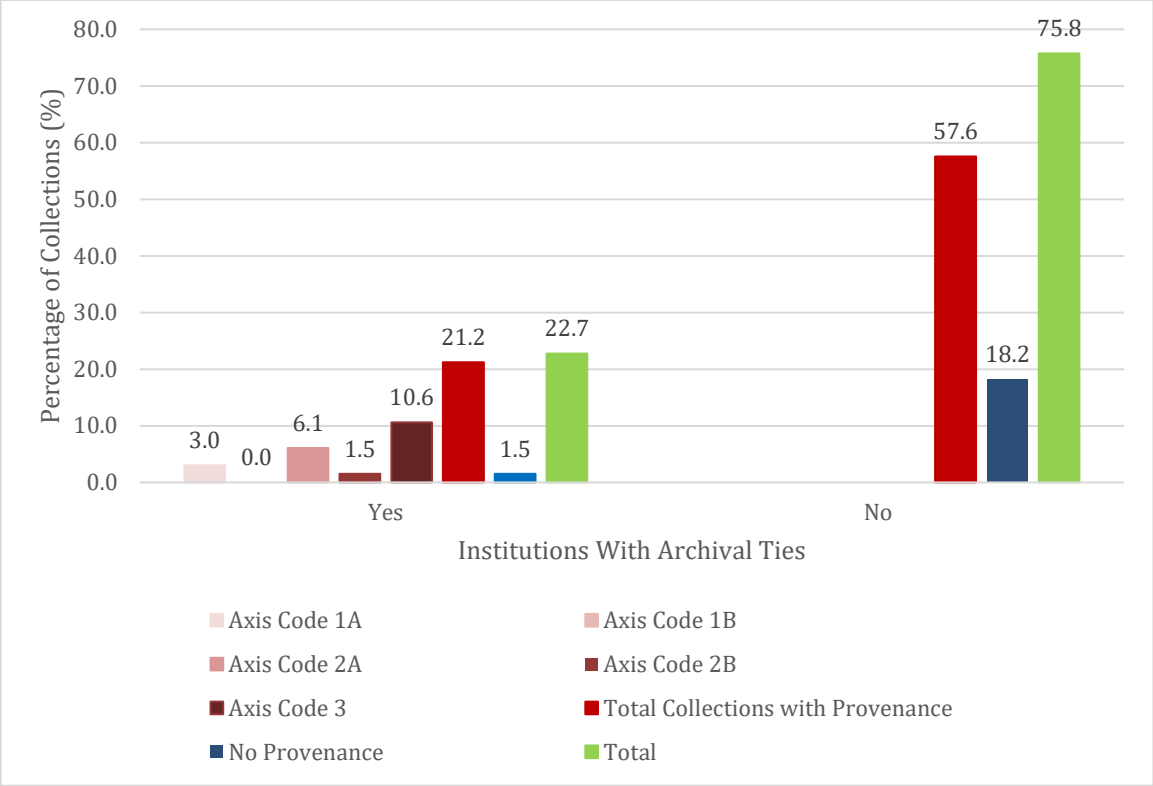

Figure C1 Bar graph showing the relation of digital special collections with ties to archival institutions and the presence of provenance metadata according to axes of provenance available within that metadata

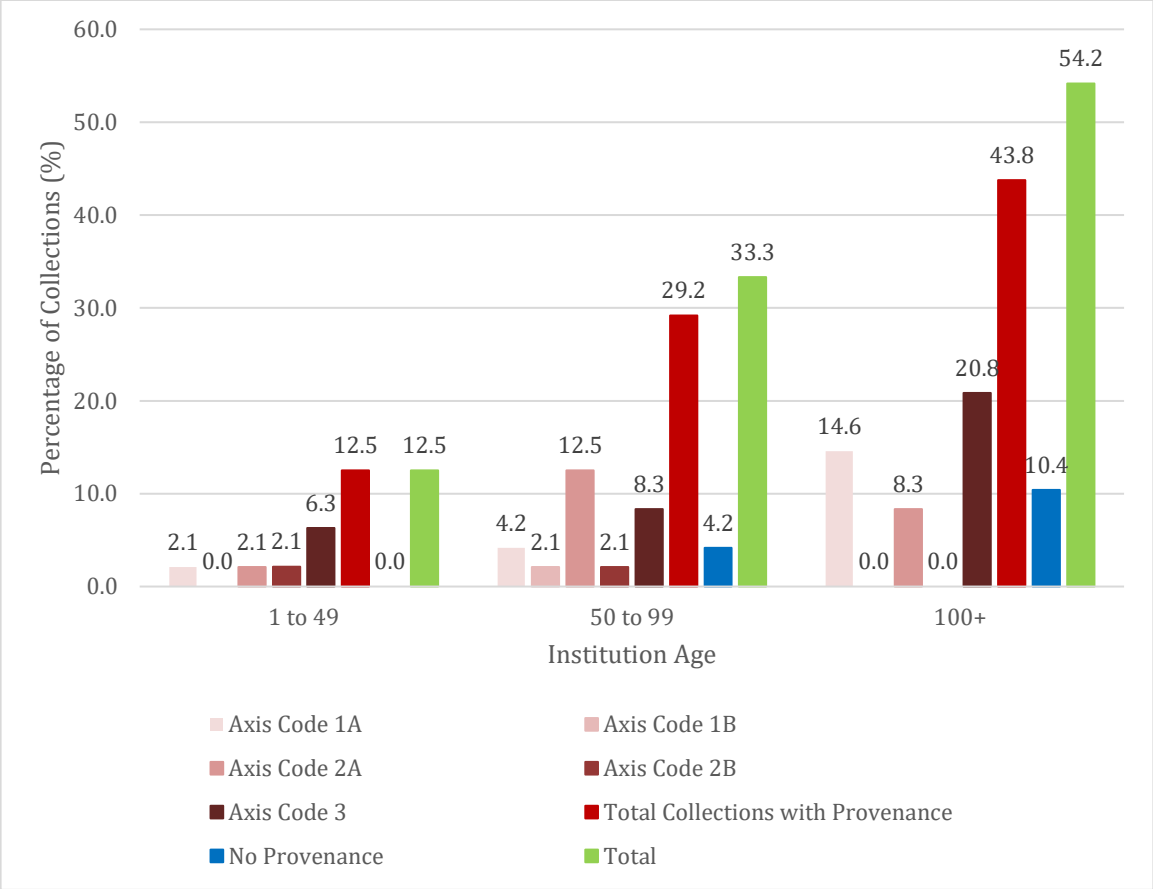

Figure C2 Bar graph showing the relationship between the age of a special collection institution and the presence of provenance metadata within the institution's digital collection according to axes of provenance available within that metadata

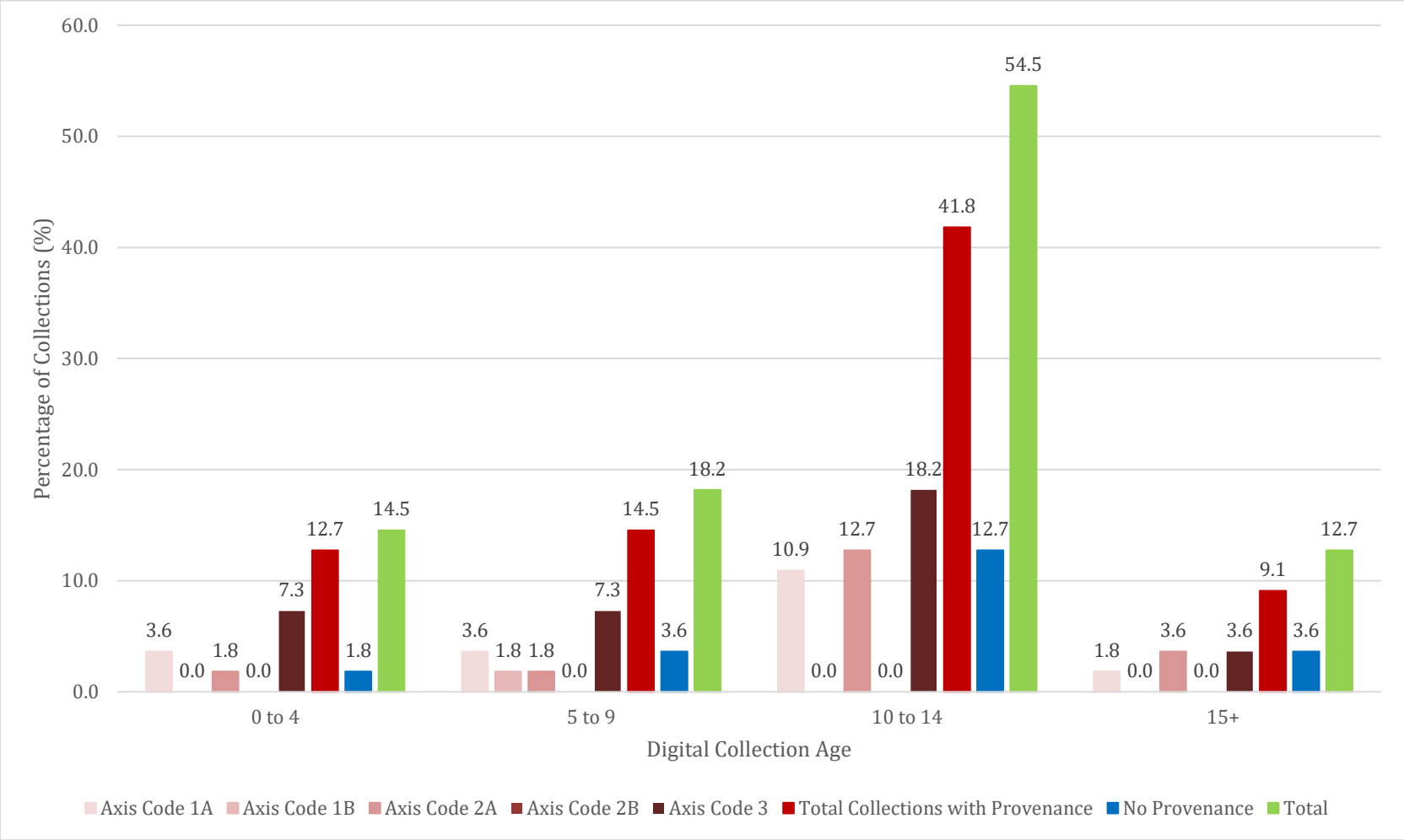

Figure C3 Bar graph showing the relationship between the age of digital special collections and the presence of provenance metadata according to the axes of provenance available within that metadata

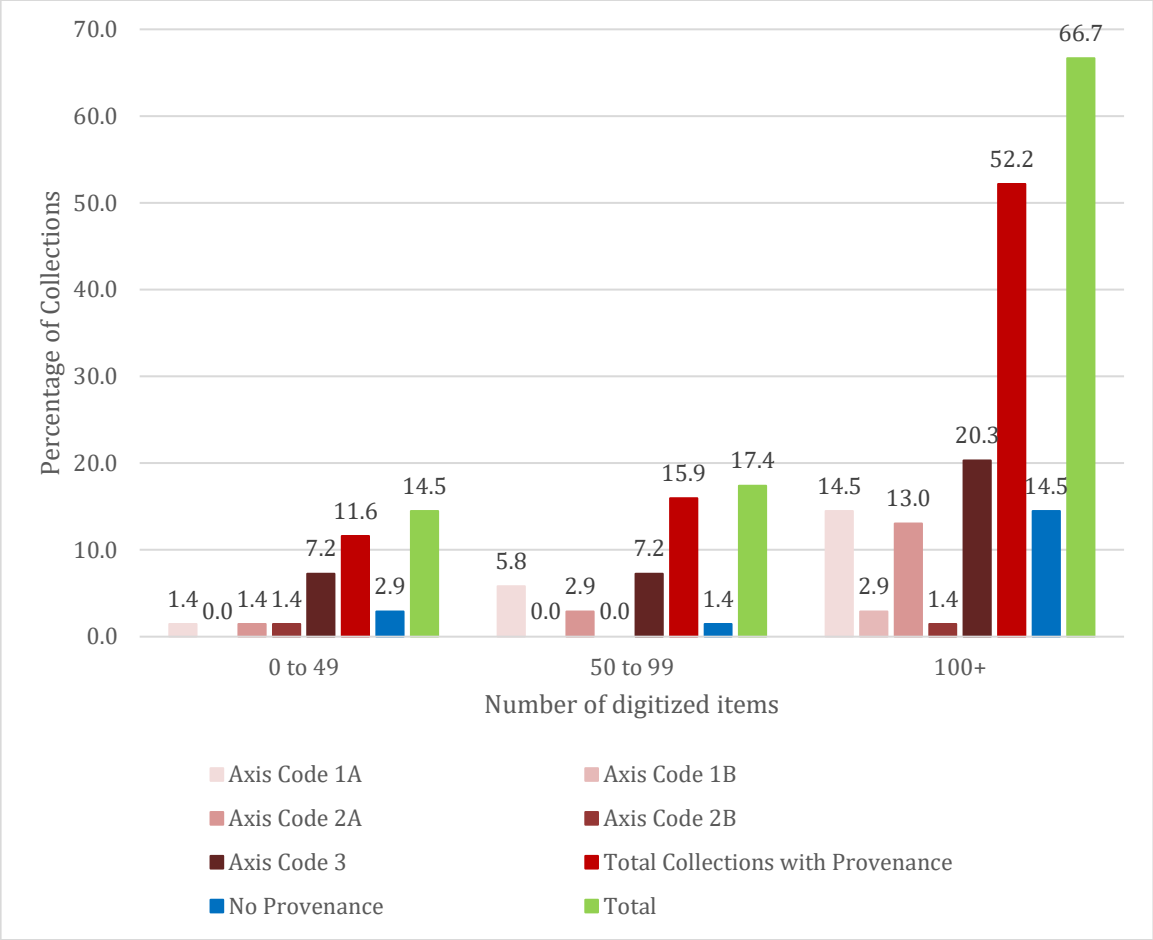

Figure C4: Bar graph showing the relationship between digital special collection size and the presence of provenance metadata according to the axes of provenance available within that metadata

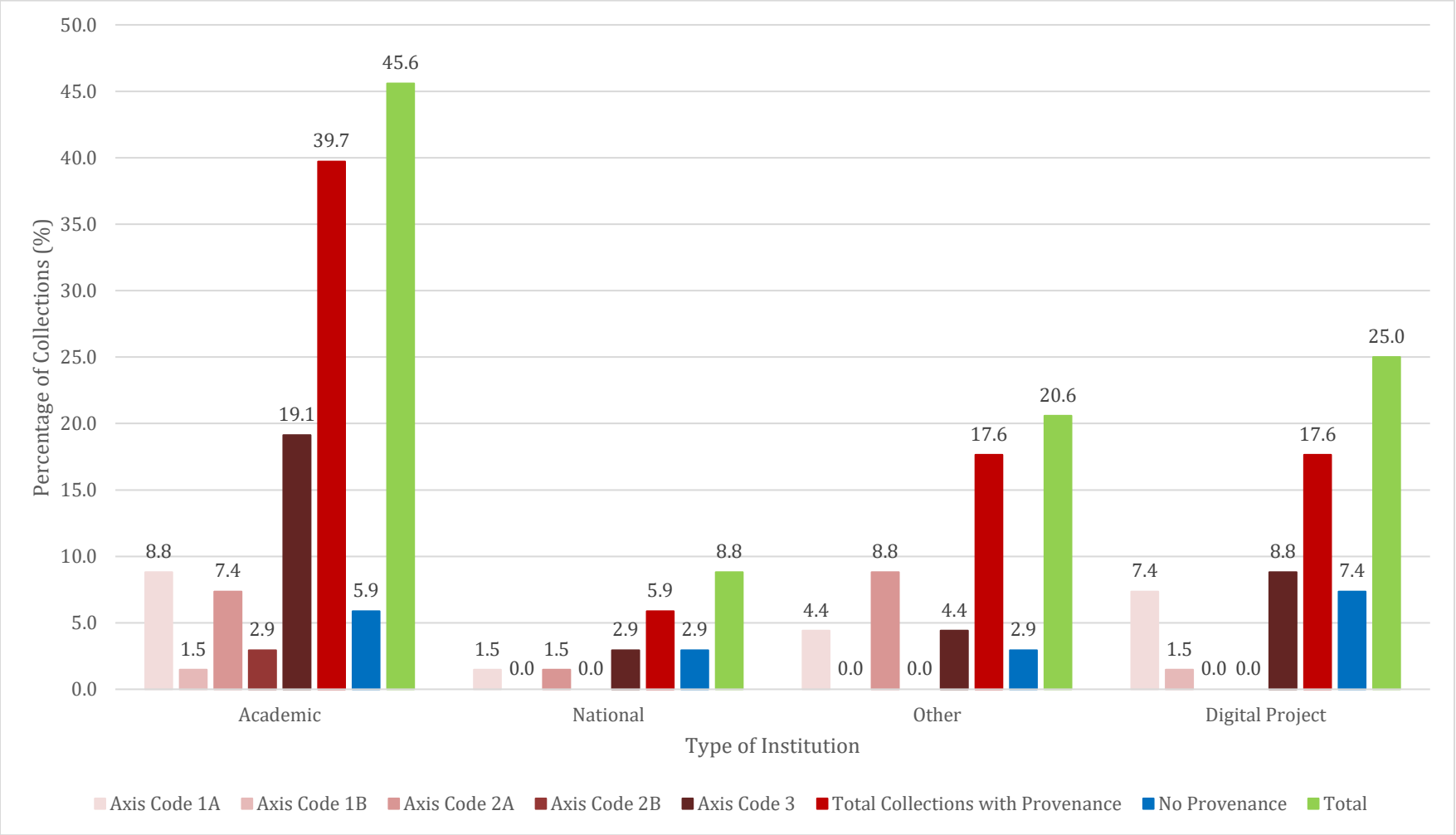

Figure C5: Bar graph showing types of institutions and their relationship with the presence of provenance metadata within digital special collections according to the axes of provenance available within that metadata

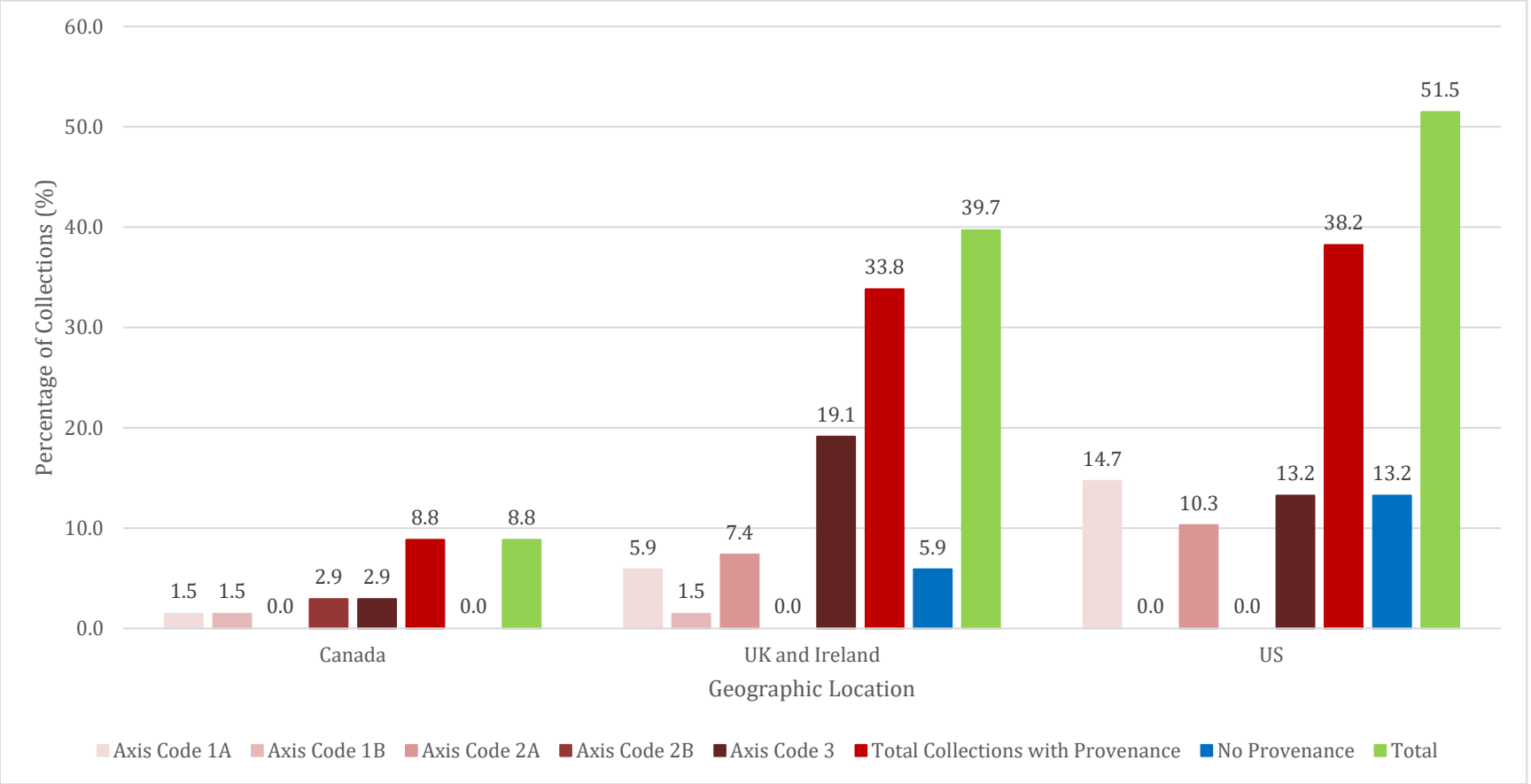

Figure C6: Bar graph showing the relationship between geographic location of an institution and the presence of provenance metadata within digital special collections according to the axes of provenance available within that metadata

Digital Features and Provenance Metadata

A number of digital features were also examined against the availability of provenance metadata and the axes of provenance available within that metadat. These included: digital platforms used by the collections, availability of linked metadata, and availability of metadata visualizations. Understanding the relationship between these factors and provenance metadata sheds light on the contemporary uses of provenance metadata (see "Provenance as Metadata") in digital special collections. Graphs for each of these factors is provided below.

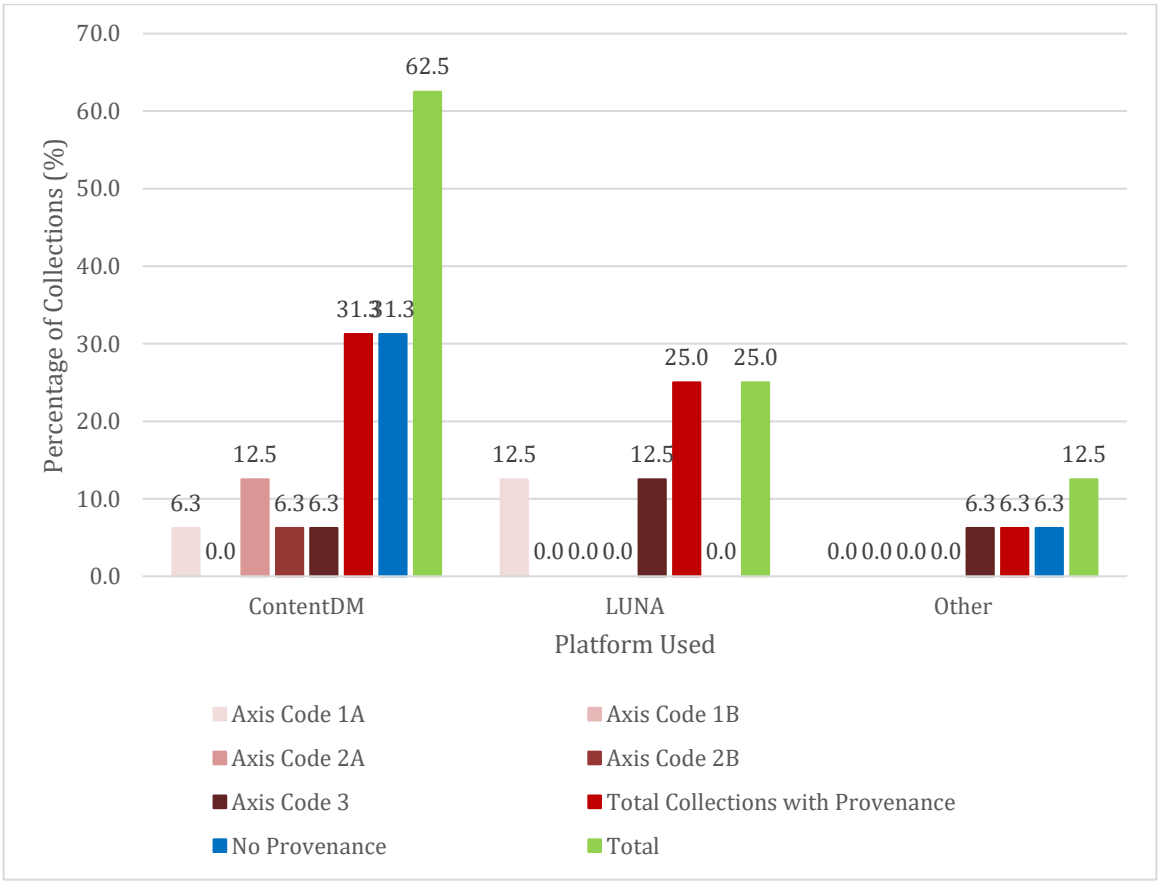

Figure C7: Bar graph showing the use of digital platforms in digital special collections and the corresponding presence of provenance metadata according to the number of axes of provenance available within that metadata

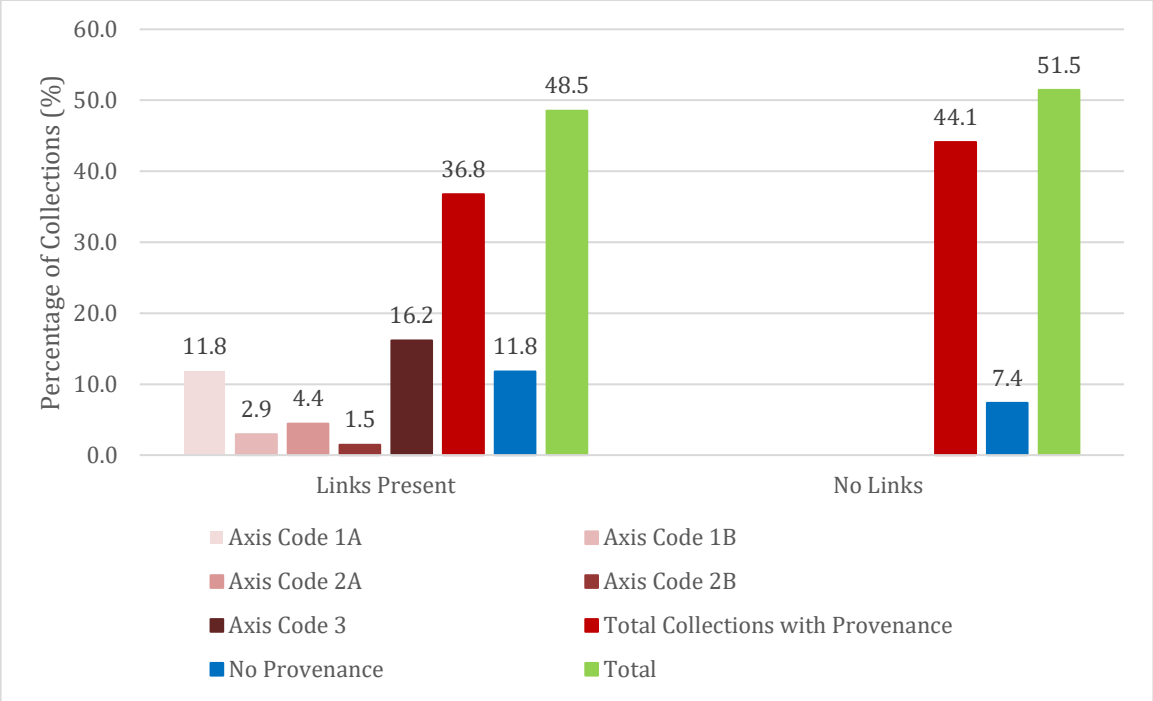

Figure C8: Bar graph showing the presence of provenance metadata in collections with and without linked metadata, according to the axes of provenance available within that metadata

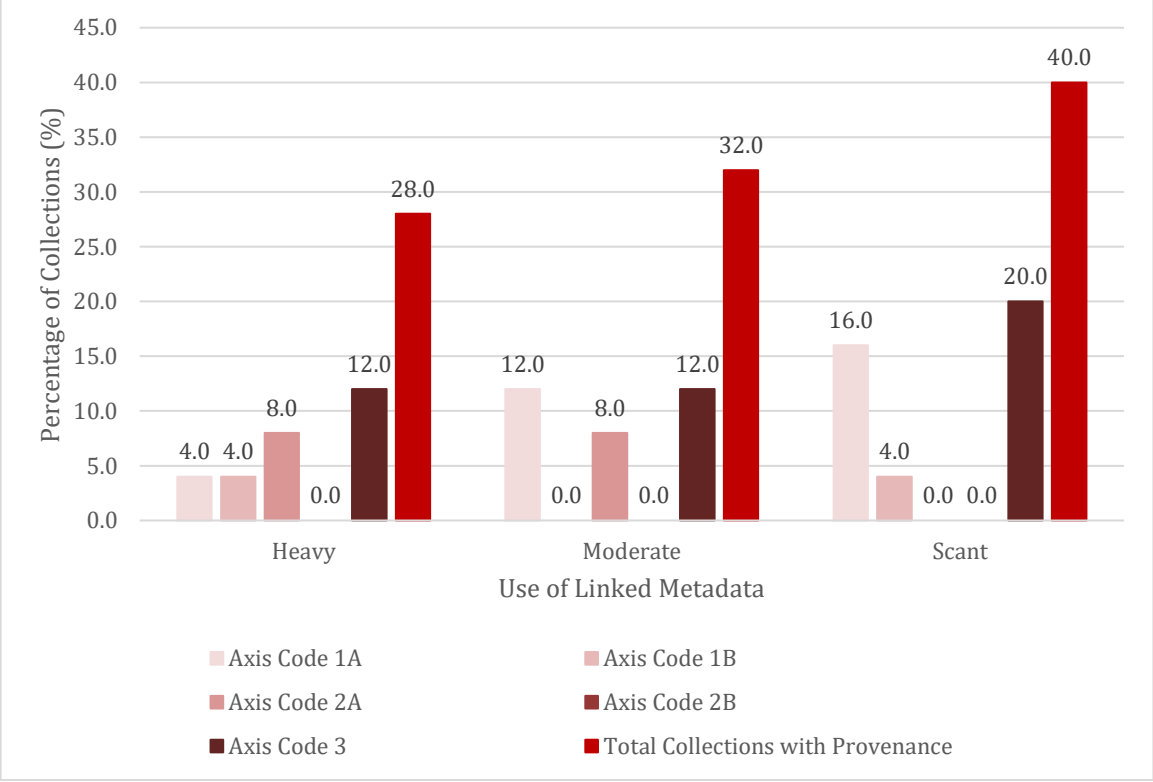

Figure C9: Bar graph showing the amount of linked metadata used against provenance metadata according to the axes available within that metadata

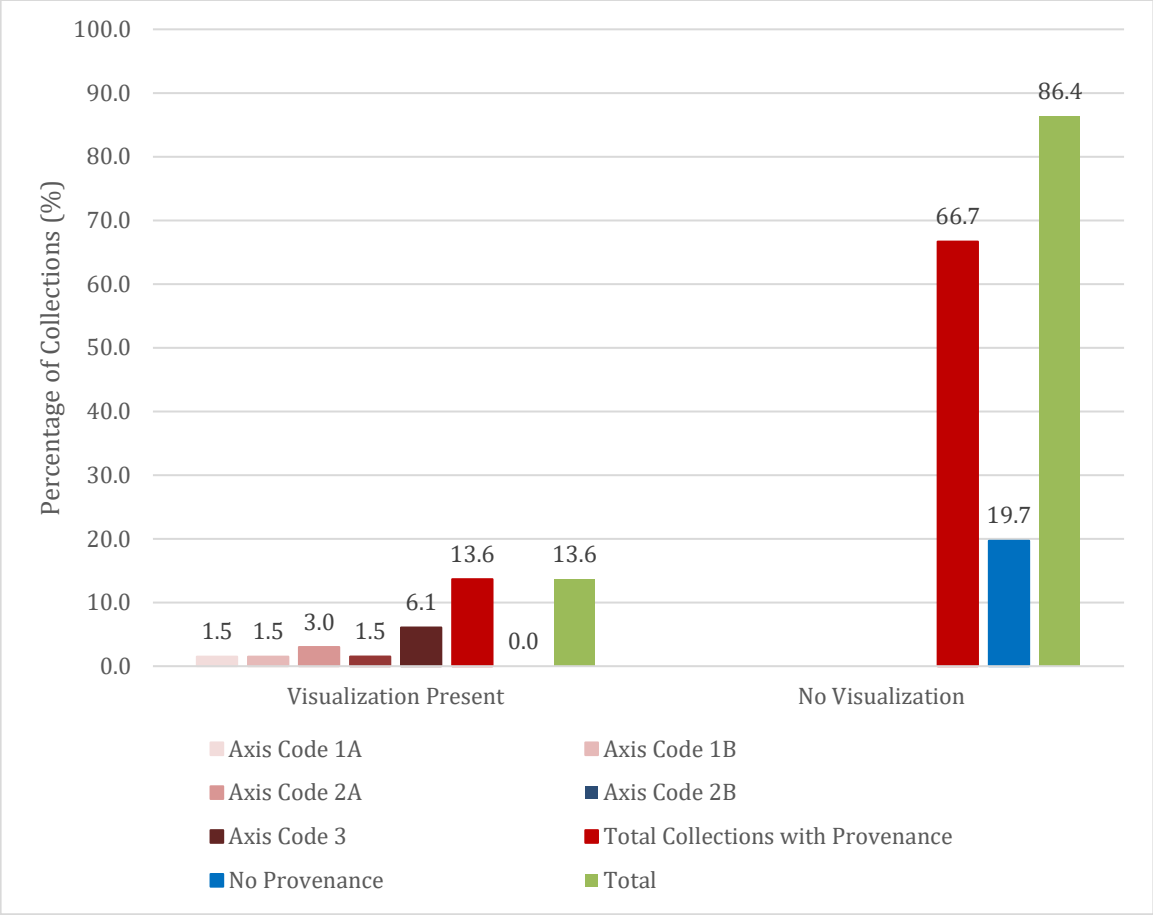

Figure C10: Bar graph showing the presence of provenance metadata where metadata visualizations were available in digital special collections according to the axes of provenance available within that metadata

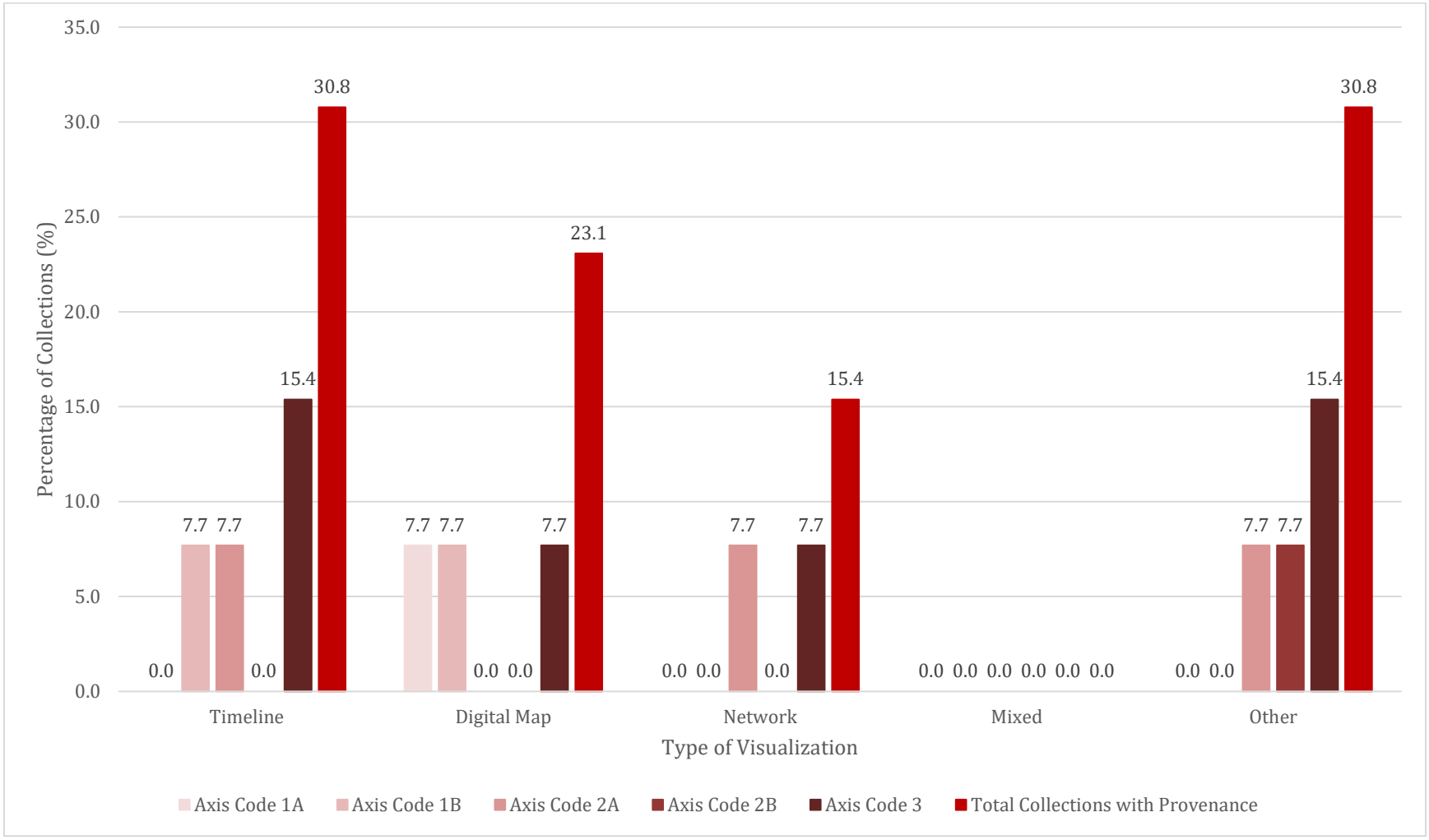

Figure C11: Bar graph showing the metaphors in visualizations available against provenance metadata according to the axes available within that metadata