Context and Digitization: Towards a New Model for Archives

Mark Vajčner

Thesis Proposal for
Doctor of Philosophy (Information Science)
Edith Cowan University

Semester 1, 2008

Principle Supervisor: Dr. Karen Anderson
Secondary Supervisor: Dr. Joanna Sassoon
Abstract

Archival digitization projects have had a tendency to be selective and to remove records from their context. Long standing archival principles have established that records have value in groups. The record’s ability to serve as evidence depends on its interrelationship to other documents and the removal from this context has negative effects on the researcher.

While researchers are dependent on the soundness of archival records, most archival research is not based on a method that considers provenance, origins, and context carefully. Researchers may not be aware of the loss of context that is rampant with digitized archival materials and the corresponding loss of evidential and research value.

The thesis will examine the centrality of theory to proper archival practice and the importance of context to historical method – as historical study is a primary use of archives. The thesis will then examine current digitization strategies, and their shortcomings, by collecting experiences of both digitization practitioners and users. The final research aim is to identify the components necessary in building a digitization project model that is both true to archival principles and that is also an effective resource for archival researchers.
Table of Contents

Abstract
Table of Contents

1.0 Introduction
1.1 Introduction
1.2 Significance and Purpose of the Study
1.3 Research Proposition
1.3.1 Core Assumption
1.3.2 Principal Question
1.3.3 Secondary Questions
1.4 Significance of the Research and Contribution to Knowledge

2.0 Literature Review
2.1 Introduction
2.1.1 Sources and Definitions
2.2 Archival Theory and Archival Writing about Digitization
2.3 Historical Method
2.4 Various Archives and History Websites
2.5 Conclusion

3.0 Research Method
3.1 Introduction
3.1.1 Historiography and Theory Building
3.1.2 The Delphi Method
3.2 Research Design
3.3 Data Collection and Analysis

4.0 Conclusion

5.0 References
1 Introduction

1.1 Introduction

Nearly a decade ago, in his inaugural address as president of the Society of American Archivists (Burckel, 1997), Nicholas Burckel painted a vision of the future in which significant scholarship would be conducted exclusively via the internet, using digitized resources. That future has already been realized in some disciplines.

Scholars in these disciplines are now discussing the impact of digital resources on their fields. The concern for authenticity and accurate digital representation are common concerns (Bearman and Trant, 1998). All researchers, both academic and amateur, have become increasingly reliant on digital information sources. Reference sources such as Wikipedia register millions of hits daily as the internet becomes the primary destination for those seeking information. Common trends in libraries are a decline in circulation of traditional paper-based materials, spectacular growth in the use of digital resources, and greater investment in digital media.

Archives are making more of their resources available in digital format, leading to investment in a vast array of archival digitization projects. Archival materials are fundamentally different from those found on Wikipedia and in most libraries. Archival materials are original records created in the course of events to document a transaction. They encompass a wide range of formats and genres. These may include paper, photographs, audio, video and computer files, recording official minutes, correspondence and diaries, to music and motion picture films. All are today being digitized and made available by archives via the internet.
1.2 Significance and Purpose of the Study

Are these resources being fully optimized? Archival digitization projects have tended to be selective of the materials they capture. This is contrary to long established principles and practices that find archival meaning in the interconnections of the whole *fonds* (or collection). Since the time of the French Revolution archival practice has been based on two key principles (Posner, 1940) – *provenance* and original order. These principles demand that archival materials be organized in such a way that their creator is known, that the records of different creators form distinct collections that cannot be intermingled, and that any organizational structure within these collections be preserved. This allows each document to remain within its context giving it greater evidential weight as it draws credibility, or discredit, from the other documents with which it is associated. Context is recognized by archivists as one of the fundamental aspects of a record (Society of American Archivists, 2005).

Yet despite the centrality of these principles to archival practice, many archival digitization projects pull materials out of their context and fail to refer researchers back to the full body of records in a convincing manner. In Canada, the Saskatchewan Council for Archives and Archivists “Saskatchewan and the Visual Arts” at http://scaa.sk.ca/gallery/art/ or the Archives of Ontario “The Archives of Ontario Remembers an Eaton’s Christmas” at http://www.archives.gov.on.ca/english/exhibits/eatons/index.html are typical examples of the display use of digital collections. These sites are not research presentations of archival collections but are more akin to an exhibition, highlighting documents that a curator has perceived as interesting or seminal.
The text that accompanies these exhibitions sometimes attempts to place the document in its historical context. This effort, however, is often less than satisfactory and rarely invites the researcher to explore the archival record more deeply. Others present more complete collections of records but fail to explain how to navigate the records effectively or to expand the research into related record sets. A good example of a complete collection, although not a strictly archival one, is Bruce Peel’s bibliography of the Canadian prairie provinces at http://peel.library.ualberta.ca/index.html, another good inter-institutional example is the “Northern Research Portal” at http://scaa.usask.ca/gallery/northern. But these provide little resources on how to use and understand archival materials. The National Archives of Australia “Documenting a Democracy” at http://www.foundingdocs.gov.au provides a good collection of documents and an excellent historical treatment but gives little in way of understanding archival collections as a whole or linking to collections from which the digitized materials are drawn.

The selection of materials for digitization is also problematic. Many projects fail to present clear selection and presentation methodologies to researchers. While selection may be conducted by individuals with a strong understanding of the topic, these selection decisions are rarely subjected to examination by an editorial board or any direct peer review process.

This lack of context may short-change the researcher. To use archival materials to their full potential the researcher needs be aware of the full breadth of records in a particular collection, which of those records relate to his or her research interest, and what other collections may offer useful information. In a traditional archive research situation the archivist would provide this knowledge, but how should this essential interaction be
replicated in the on-line world? Few digital presentations of archival materials have considered this question. Archives are missing a strong opportunity to use the medium to educate their users about archives and about how archival materials are organized and effectively used. The internet affords a perfect opportunity for this as evidenced by such digitization projects as that of the Archives Society of Alberta. “Archives in the Classroom: Letters from the Trunk” at http://www.ataoc.ca/archives/main.html is an excellent example of an online archival learning resource that was designed jointly by archivists and educators for use in the classroom. Another useful example is the website of the John Curtin Prime Ministerial Library (JCPML) at http://john.curtin.edu.au/. JCPML offers a range of education tools including onsite, online, and outreach programs targeting students from primary school to university.

Thus there are several key dichotomies evident in the current state of archival digitization. The central dichotomy is the relationship, or lack thereof, between archival theory and the practice of archival digitization. Another key dichotomy is the relationship between archives and researchers. Particularly important to this dichotomy is the understanding of archives by researchers and the understanding of research needs by archivists.

1.3 Research Proposition

1.3.1 Core Assumption

It is assumed that for archival records to truly be useful, they must meet criteria to serve as evidence of an action. It is also assumed that the research methodology of most archival users is not archival based. Or, to put it another way, research is not based on a method that considers provenance,
origins, and context. A rigorous process of selection, organization, and display needs to evolve to make archival digitization efforts more useful to the researcher, both academic and amateur. This process needs to be firmly grounded in existing archival theory to ensure a presentation of material that serves as evidence of past actions.

1.3.2 Principal Question

The principal research question is

Why is archival theory important in archival digitization?

Theory is the central pivot around which archival practice moves. This is not the case with archival digitization. There is a need to show that theory is important to practice in this archival endeavor as well. From this principal question, several secondary research questions arise.

1.3.3 Secondary Questions

What archival theories, and archival practices built on those theories, are important to archival digitization projects if those projects are to respect central archival principles of provenance, original order, and context?

There exists today a dichotomy between current digitization practice and accepted archival theory. The criteria of archival research value should be the same for resources available digitally as well as in more traditional formats. Subsidiary questions that may be asked at this point include:

Why is there currently a divergence between archival theory and digitization practice?

Have there been similar divergences between theory and new practice/methods/technologies in the past? How were these resolved?

And
How should the divergence be resolved now?

In fact archives have just recently resolved a divergence between theory and practice. During the 1970s, as electronic records became prevalent in our society, many archivists argued for an abandonment of archival theory. They believed that computer generated records were so fundamentally different that archives could not possibly respond to the challenge of archiving them in traditional ways. But since the 1990s electronic records archivists have re-discovered archival theory, applied it thoughtfully to computer generated records, and have developed archiving models that respect provenance, original order, and context. Our current practice with archival digitization may mirror this experience.

These questions constitute the theoretical component of the research proposal. Further questions will be asked regarding dichotomy between the understanding of archives by researchers and the understanding of research needs by archivists.

Here archivists and archival researchers will be extensively consulted during thesis research. This secondary question will focus around the needs of the archival researcher and how both the archivist and researcher envision them.

The question is

What is the current conception among archivists of an ideal digital resource website? And what is the conception among archival researchers?

As already mentioned, archival meaning is found by the interrelationship of documents. For example, does the memo corroborate what is written in an
earlier report? Does the letter provide an alternative view of the event described in the memo? Which version of events is most trustworthy? These are the questions asked by the serious researcher of archival materials. So

*Is authenticity important to the archival researcher?*

*Does the archival researcher have an understanding of why proven authenticity may be important to research?*

And, with regard to the authenticity question,

*What components would a useful website have?*

In the end, all these questions will be drawn together to answer the following overarching question

*What components are necessary in building a digitization project that is both true to archival principles and is also an effective resource to archival researchers?*

### 1.4 Significance of the Research and Contribution to Knowledge

As archival digital resources become increasingly prevalent, clearer thinking about digitization practice will become crucial. While the technical cost of digitization declines as scanners and computer memory continually fall in price, the organizational costs remain constant. Materials cannot simply be scanned. Much more is required of a digitization project. Proper projects require planning, quality checking, and considerable organization to be useful and effective. As the technical costs decline and the organizational expense becomes an ever increasing part of the whole, it is only natural that more attention be paid to how that organization can be optimized.

It is essential that archivists contribute their professional expertise to this discussion. As we enter a world where the only interaction between
researcher and archivist may be virtual, it is incumbent on archivists to develop rigorous digitization practices that take account of archival theoretical traditions and user needs.
2 Literature Review

2.1 Introduction

This literature review encompasses three main areas. The first section focuses on archival theory and on archival writing regarding digitization. The second section explores writing on historical method, or how historical research is conducted and how history is written. Historians are major users of archives. Archival theory and historical method have developed in tandem and remain closely tied. The third and final section examines archival and history websites. These will be important in ascertaining exactly the current state of digitization practice.

2.1.1 Sources and Definitions

Several sources were used in the research for this literature review. Much archival writing is published in journals and relatively few book-length monographs are written. The major English-language archival journals have been extensively searched for materials relating to digitization. These journals are *American Archivist*, *Archival Issues*, *Archival Science*, *Archivaria*, *Archives*, and *Archives and Manuscripts*.

In most instances titles included have been limited to digitization, that is, the production and management of electronic copies created from physical materials already held in archival repositories. These copies may have been created for access, preservation, or publicity purposes. Generally, this literature review does not include titles that concern themselves with the management of electronic materials created by external agencies that are
subsequently acquired by archives. These materials face unique challenges that are beyond the purview of this literature review and this proposed study.

A further restriction has been to limit titles to theoretical treatments of digitization rather than include progress reports on a particular project. The literature is full of the latter; these, however, provide little empirical material for this literature review.

For the purpose of this study archival digitization is defined as the making of electronic copies of archival materials and their presentation on the internet to aid researchers and increase access to historical documents. The original archival materials may be textual, photographic, or recorded sound and images. By digitization it is meant that these ‘originals’ are copied into a format that makes their distribution over the internet possible. The intent of that digitization is neither preservation nor replacement of the original. The intent is to provide researchers with an alternate means of accessing the information contained in the original archival document.

This study will focus primarily on digitization conducted by archives, or formal institutions specifically established to manage documents in accordance with established archival principles. It is recognized that other institutions, and even individuals, engage in digitizing historical documents. These include libraries, museums, art galleries, historians – both professional and amateur, history departments, genealogists and genealogical societies, and historical societies among others. In some cases the digitized collections and processes of these others will be examined and commented on but the focus of this study is to postulate on a model of digitization intended for archives in the narrower sense. Libraries, museums, and others are less likely
to apply archival theory in their handling of physical materials, let alone in their digitization projects. Should a method of digital presentation emerge in archives, that takes into account the expertise that archival theory provides, this may become the standard for all scholarly research in digitized archival materials no matter what agency digitizes them.

2.2 Archival Theory and Archival Writing about Digitization

In her 1981 essay, “The Originality of the Avant-Garde”, Rosalind Krauss stressed the importance of evidential value. “The theme of originality,” she wrote, “encompassing... the notions of authenticity, originals, and origins, is the shared discursive practice of the museum, the historian, and the maker of art. And throughout the nineteenth century all of these institutions were concerted, together, to find the mark, the warrant, the certification of the original” (Krauss quoted in Koltun, 1999, pp. 134-35, Original in Krauss, 1981, p. 58). For archives this has remained a primary goal of theory and practice. In today’s world the digitization of archival collections poses a specific challenge to archives and to this goal of ensuring authenticity.

Archives have always served an evidential purpose. A background to the theoretical underpinnings of archival theory can be found in numerous writings. The history of archives and the development of archival thinking have been well researched and explored. Duranti (Duranti, 1993) traces that history back to the dawn of civilization and the keepers of the cuneiform tablets of Mesopotamia. Posner (Posner, 1972) does as well, but places the development of modern theoretical concepts of archives with the French Revolution (Posner, 1940). A synthesis of these and others (d’Addario, 1992; Cook, 1997; Higgs, 1996; Ketelaar, 1996; Posner, 1967; and Rumschöttel, 2001) outlines the central role of records as evidence in archival thinking. This
synthesis will be important in establishing the theoretical values of archives which managers of digital projects must keep in mind. Briefly stated, in order to act as evidence, archival documents must be considered not as individual items but as part of the whole collection. The interrelationships within these collections, or the context of the individual record, determine its authenticity and value as evidence.

It is important to consider what context means, what it has meant, and how it’s meaning has evolved. These questions are the crux of the matters that this thesis literature review will examine. They will be explored and developed to consider what types of contextual information is relevant to research today. For example, do different types of researchers need different types of contextual information? If so, what types and why? The digitization model that will be proposed by this thesis must ultimately answer the following question to be effective: how should contextual information be made available online?

Evidential value is a central concept in archives yet it is a difficult concept to define. Understanding of evidential value by archivists suffers from the assumption that the concept is self-evident and efforts at definition have tended to be narrow legalistic definitions. Archival theorists have worked to clarify archival ideas of evidence but that clarification is still a work in progress. The thesis literature review will draw on these recent efforts to define the concept (Cook, 1997b; Brothman, 2002; Meehan, 2006). The meanings of other basic archival concepts are not straightforward or uncontested either. Some had definite meanings that the change of prevalent theory has made less so. Like the issue of context explored above – meaning has evolved. These definitions and changes will be explored further in the
thesis and include context – both physical and historical, provenance, and origins.

Archivists need to focus on questions of context. In the 1970s electronic records posed a conundrum for archivists. How could ever-changing database records be considered in the same light as stable paper records? At one time in the approach to dealing with electronic records traditional archival theory was considered to be outdated but then archivists rediscovered their base theories, applied them thoughtfully to the new media, and came out with a stronger understanding of what made archives and what constituted a record (Cook, 1997, pp. 40-43, Nesmith, 1993, pp. 2-28). Hopefully, at some point in the not to distant future, archival experiences with digitization will similarly develop a stronger understanding of the continuing importance of context. The research undertaken here will aim to contribute to this shift in thinking about digitization from simply a raw response of archives to technology to a more considered thinking about the possibilities of digitization.

There is a critique of this role for archival theory (Brothman, 2002) that questions the concept that evidence is the governing purpose of archives. Its main contention is that late-twentieth century organizational and technological cultures make it extremely difficult to create records that are evidential. Furthermore, current discussions of evidence are blind to the contradictions raised by these contemporary cultures. Nevertheless the application of archival theory to the management of electronic records has resulted in more successful electronic archiving than in the days when archival theory was ignored.
In some ways archivists are beginning to consider the loss of context and the resulting effect on the record. Joanna Sassoon (Sassoon, 1998) examined the effects of the loss of context on photographs. Photographic theory has identified the materiality of the print and negative as providing a part of the overall meaning of the photograph. Sassoon argues that these material features are precisely those that are lost during digitization and that this loss has an adverse effect on the interpretation and meaning of the photograph. Others go further, claiming that the materiality of the record can communicate meaning between generations and cultures (Rekrut, 2005). In this case the record, as object, offers researchers a personal and direct sensory connection to the past and this physical experience can be recognized as a primary source of context and evidence.

Archives have always had an affinity for treating their documents as objects and the physical composition of records continues to hold an important place in archival theory. As the nature of records changes and as their physicality takes on a more ephemeral nature, archival writing is considering a rethinking of this ‘object’ oriented view (Rylance, 2007). The physical context of records, however, remains important as so much of the archival record continues to have a physical presence. This physical context will pose a problem to archival digitization that may be irresolvable.

The publications resulting from the three I-CHORA (International Conference on the History of Records and Archives) conferences are an important source in understanding this trend and others in recent archival thinking. Archival theorists such as Terry Cook, Richard J. Cox, Eric Ketelaar, and Tom Nesmith among others presented and later published papers concerning the archival record and context in the modern world. Randall C.
Jimerson, and Margaret Procter did the same on the relationship between archivists and historians. These publications, appearing as special issues of Archivaria, Archival Science, and Libraries and the Cultural Record will be included and considered in the thesis literature review.

Many contemporary archival theorists have begun to consider what Terry Cook has called a “paradigm shift” in archival thinking. This shift is away from the empirical positivism of Sir Hilary Jenkinson and Theodore R. Schellenberg towards a more postmodern view of both the profession and records.

Expressed succinctly, this view asserts that we are in a postmodernist era and in this era we must recognize the postmodern distrust of the modern. The belief in universal truth, or an objective knowledge based on principles of scientific rationalism, no longer holds sway. In our postmodern world nothing is objective or impartial and everything is shaped by its speaker for a set purpose (Cook, 2001, pp. 5-7). No document or record can be considered an innocent product of action but rather is a “carefully” and “consciously” constructed product.

As Cook writes (Cook, 2001, p. 7), “Texts (which include images) are all a form of narration more concerned with building consistency and harmony for the author, enhancing position and ego, conforming to organization norms and rhetorical discourse patterns, than they are evidence of acts and facts, or juridical or legal frameworks”.

This is particularly true in the digital world where documents may be altered, broken apart, and recombined with simple ease. Lilly Koltun
describes the use of archival clips in film and the ease with which meaning may be constructed. An obvious example is the central character of *Forrest Gump*, who is shown meeting three American presidents in various altered news clips. These are obvious fabrications and thus possibly harmless. Others may not be. In Oliver Stone’s 1994 film, *Natural Born Killers*, a dream sequence utilizes archival footage of a Native canoe with a shaman in bird costume. This genuine archival sequence was itself orchestrated for a 1916 melodrama and loosely represents Native custom from a part of British Columbia. The implication of Stone’s including it, to represent a desert shaman of New Mexico, Koltun contends, is to suggest unconsciously that all Native people are to be understood as similar (Koltun, 1999, pp. 126-27). Such use of documents, to represent utterly unconnected concepts or meanings, will be more prevalent in the digital age. This is central to this part of Koltun’s argument (Koltun, 1999, p. 126), that the digital record and the digital copy encompasses an enhanced potential for mutability, for easy re-contextualization... and re-construction. This ease is integral to digitization’s very nature, and means we now need to ask more urgently how intrusive archivists will be, and allow others to be, in use and manipulation.

The consideration of what exactly is an authentic digital document is beginning to occupy archival thinkers. A decade or two ago, the primary concern with authenticity was in assuring a complete and accurate transfer of electronic records to archives from depositing agencies (Acland, 1991; Bearman, 1994; Cook, 1992; Dollar, 1992; Waters and Nagelhout, 1995; Wallace, 1993). Today greater concern is placed on the authenticity of existing digital documents for research (Bearman and Trant, 1998). Concern centers on ensuring that the object of study is in fact exactly what it purports to be.
This, to a certain degree, runs contrary to the traditional archival vision of the role of archivist. The profession has historically not viewed itself as gatekeeper nor as interpreter. Archivists acquire the record, aim to ensure its completeness, organize in a manner not to threaten evidential value, and then leave the record to speak for itself. As David Bearman and Jennifer Trant write, interpretation and re-interpretation of primary and secondary sources is “the foundation of much humanistic scholarship”. Furthermore the

Construction of a convincing argument depends on an evaluation of the authenticity of source materials. Judgments about authenticity are based on assessments of the origins, completeness and internal integrity of a document. (Bearman and Trant, 1998)

These judgments have traditionally been left to the scholar, who, archivists have assumed, would rigorously judge the record. Will the filmmaker also follow such a method, or even care to? As early as 1984 studies had shown that what archivists have tended to view as their primary clientele - scholars - were being eclipsed by other users less interested in exhaustive primary research (Freeman, 1984). Their understanding of context, and the need for it, was limited. Limited not because of any intellectual deficiency, but limited because they were not necessarily seeking the record for evidential purposes. Digitization is bound to increase the distance between the users and archivist as archival materials become more readily available. The archivist must step up to the plate for context.

Unfortunately there is no agreement among archivists on what course should be followed. In his 1999 presidential address to the Society of American Archivists, H. Thomas Hickerson presented ten challenges that he saw facing the archival profession. Number five on his list was accessibility. Digitization, he felt, was the primary method to achieve this goal (Hickerson,
2001, p. 11). He felt, however, since context wasn’t the primary concern of researchers it shouldn’t necessarily be the focus of those archivists making records available digitally. “In developing digital resources,” Hickerson writes

we initially focused on content, but we are now beginning to turn our attention to designing and implementing the services necessary to support effective collection use in a networked environment. The use of focus groups and the analysis of data recording patterns of on-line navigation and use will help us develop useful and efficient means to meet the service needs of our various clientele.

Focusing on our users implies that we acknowledge the primacy of their needs and respond by utilizing methods that address those needs. (Hickerson, 2001, p. 11)

Hickerson goes on to cite a study that analysed user requests for photographs from the University of North Carolina at Chapel Hill and from the North Carolina State Archives (Collins, 1998). It concluded that the overwhelming majority of requests for photographs were not for historical research but for some other purpose and that not a single researcher requested to see an image in the context of its collection. “I do not need to comment” Hickerson writes (Hickerson, 2001, p. 11), “on the meaning of these findings regarding the adequacy of access approaches that are largely provenance-based”.

Of course these researchers in North Carolina took for granted that they were visiting established and reputable institutions where the authenticity of the records was affirmed. Hickerson also takes for granted that the use of contextualization in the organization and keeping of materials can allow archivists to use other principles in their access systems. These important questions may be taken for granted in the digital world where
anyone can post and publish information. Many researchers today are information-savvy and can use Google to find just about anything. Fewer are knowledge-savvy and find it difficult to test the information they find. Archivists have an increasingly important skill at their command.

Archival writing on enhanced online descriptive systems includes Michelle Light and Tom Hyry (Light and Hyry, 2002), Heather MacNeil (MacNeil, 2005), and Tom Nesmith (Nesmith, 2005). Ian Wilson (Wilson, 2007), national archivist of Canada and a strong advocate of digitization considers the importance of public programming in increasing the awareness of archival materials. The work of Michael Moss, and other theorists at the University of Glasgow, and Margaret Hedstrom’s work, at the University of Michigan, on archival interfaces with users in the digital age is noteworthy. All these theorists will be included and considered in the thesis literature review.

Bearman and Trant believe that even the seasoned scholarly researcher is facing difficulties in the digital environment. Forgeries and fakes have long been a concern of scholars in the humanities. In the past the technical barriers to making a plausible forgery were formidable. As a result of digital technologies, that is no longer the case. Furthermore, works were scrutinized and authenticated before entering “an authoritative information stream” (ie: archives, libraries, museums, art galleries, or published in peer-reviewed journals, books, etc.). Today scholarly publishing over the internet is still very much in its infancy. Publishers are experimenting with the medium attempting to replicate traditional publishing and peer-review models on the internet. At the same time new models are emerging, such as the open source movement.
Authentication models are also in their infancy. Bearman and Trant propose a host of solutions ranging from digital watermarking and digital signatures to the creation of certified deposit “collections or record”. They argue for developing metadata structures to carry authentication information. These are really new terms for what are established concepts. Archives have been and continue to be “collections of record” and authentication is an important part of archival theory. That Bearman and Trant must call for the creation of those tools indicates that archives have not been exceedingly successful in communicating their expertise in understanding context.

That understanding of archival expertise may come from other professions. In her 2000 report to the Council on Library and Information Resources, Anne Gilliland-Swateland argued that information professionals need to study each other to effectively respond to the challenges of the digital world. Abby Smith, Director of Programs for the Council wrote in her preface to Gilliland-Swateland’s report, “librarians often seek answers to questions that archivists… have dealt with for years” (Gilliland-Swateland, 2000, p. iv).

Speaking to the Australian Society of Archivists at its annual conference in 2000, Elizabeth Hallam Smith declared that many archivists today are enthusiastic to embrace digital technology and the internet as a channel for service delivery (Smith, 2000). But she saw several important questions before archivists, among these: How much do we know about the ways in which archive users access information? And to what extent do archives impact on the consciousness of non-users with a general interest in matters cultural?
Smith cited studies that showed that 51.5% of archive users in the United Kingdom in late-1999 had Internet access. Another PRO Reader Survey, conducted in the following year, produced a higher figure still, of 67% (Smith, 2000, p.2). Significantly, these archives had no idea as to what researchers were using and how. There is recent research on library user information seeking behaviors, but in the case of archives we are in the dark, since there is no academic research on such behavior in archives.

Elizabeth Yakel and Deborah Torres (Yakel and Torres, 2003) argue that archivists have never defined the characteristics that denote, in the eyes of archivists, an expert archival user. Their 2003 study, involving in depth interviews, led to the development of a model of researcher expertise for use in archival user education programs. Yakel and Torres argue that there are three distinct forms of knowledge required to work effectively with primary sources. These are: strong subject knowledge; artifactual literacy; and what Yakel and Torres call “archival intelligence”. Archival intelligence is a researcher’s knowledge of archival principles, practices, and institutions, such as the reason underlying archival rules and procedures, the means for developing search strategies to explore research questions, and an understanding of the relationship between primary sources and their surrogates. At about the same time Wendy Duff and Catherine Johnson (Duff and Johnson, 2003) focused on genealogy users. Their findings provide information on the stages of genealogical research, how genealogists search for information, the access tools they use, the knowledge required, and the barriers they face.
In truth archivists are still hazy about exactly who their users are – either digital or in person. As early as the 1980s Elsa Freeman showed that the majority of archival users were not academic historians, as archivists were wont to believe, but were drawn from a wide variety of fields using archives for a multitude of purposes (Freeman, 1984). A 1999 study of archives in the United Kingdom concluded that the educational sector and professional researchers made up only 16% of archival users while fully three-quarters of research in archives was a result of personal interest, the most significant of which was genealogy and family history (Smith, 2000, pp. 4-5).

While these studies show that archivists are not clear as to who their researchers are, it can be said that the nature of the research itself is clear. The nature of the research is in someway historical. Freeman shows that academics may be outnumbered by other researchers, but those researchers, like the academic historian, are interested in discovering something about the past. The same is true of the figures quoted by Smith – genealogy and family history are a form of historical study. Thus an understanding of methods of historical enquiry is important for archivists and for archival digitization design.

2.3 Historical Method

Archival theory has, in many ways, developed in tandem with historical method. Modern archives and modern historical method both date from 19th century Europe and post-modernist thinking has impacted both within the last thirty years. This close relationship is the result of the similarity of their subject – both disciplines deal with the past – and the similarity of the individuals who inhabit both fields – those with a desire to
understand and animate the past. Historians, for example, have played a influential role in the shape and direction of archival thinking.

The practice and method of history was revolutionized in the early part of the 19th century. Until then history had been considered a branch of literature – a story of the past used to teach morals and tactics. In an 1831 letter to his brother, Leopold von Ranke (Taylor, 1967, p. 114-15), a pioneer of modern historical method, outlined the new view of history. “My basic thought”, Ranke wrote, “is not to accept one theory of history or another... but to recognize the facts, to master them and display them”. These facts were to be found in the documents and the consequence of Ranke’s doctrine was that if the archives were exhaustively utilized, history, as it actually was, would be revealed. Ranke, time and time again, scoured the archives looking for that one document that contained the facts that would allow everything to fall into place and history to be revealed.

The historian was no longer a writer but a compiler of the evidence. The new method of diplomatics arose to study the document and its text to determine its authenticity and thus its value as evidence. The historian, it was argued at the close of the 19th century, was now a scientist because a detailed methodology to analyse the evidence had been developed.

More modern historical methodology, as defined by Susan Grigg and Stephen Humphreys has, in some respects, changed little since the days of Ranke. Both Grigg and Humphreys see a firm footing in the archival evidence. Historians, whether their field be political, economic, social, or cultural will generally concede the importance of primary sources (Grigg, 1991, p. 229). These sources become the raw materials of the historian’s craft
and the relationship between the historian’s thought and these sources – the evidence of past actions – is the central question of historical method.

Humphreys writes (Humphreys, 1980, p.8):

The exact reproduction of documents may not seem a very creative or inclusive intellectual activity, and certainly it constitutes at best only a very small proportion of what historians actually try to do. But even so, it is sufficient to assure us that there is an important existing and knowable reality to which historical description can be directly linked. This in turn gives us a true objective foundation for historical thought.

The analysis of historical resources – its method and proper process – has occupied incalculable pages in scholarly journals. But research, and how research is done, has never been strictly defined by the profession (Grigg, 1991, p. 228). The methodology of Ranke compared research to the physical act of mining. The historian sifted through the available evidence until that one “nugget” of a document was found that allowed the subject of study to be fully revealed and all elements to fall into place. It was an almost providential revelation that suited the Protestant Ranke and other late-19th century historians.

Another 19th century German, Ernst Bernheim, made an effort to define evidence and divided primary sources according to the purpose of their creation (Grigg, 1991, p. 229). Did the author intend to convey an interpretation to the reader? If so, the evidence was less reliable than a document that simply recorded an observation. This second example was a document solely of the facts from which the reader or researcher could draw his or her own conclusions. This search for a hidden agenda in the document became the quest of the 20th century historian. When Herbert Butterfield replaced Providence with the interplay of interests as the driving force of history in the 1930s – what he called its “clash of wills” – the scepticism about
the document, or what interpretation the author intended to convey, became an implicit part of historical study. If history became a “clash” then the document was simply another tool in advancing the ‘battle’ and naturally contained a meaning designed to aid or defend its author in his or her continuing struggle of wills. It led G.J. Renier to write, in his seminal historical method manual, *History: Its Purpose and Method* that historical description was based not on “sources” but ‘traces” gleaned from those documents and carefully evaluated and scrutinized by the historian using his or her expert knowledge of the past (Renier, 1950. Quoted in Humpherys, 1980, p. 3). European historians had long taken the view that primary sources were essentially traces of historical activity (Grigg, 1991, p. 231). Collingwood wrote of them in his *Idea of History* as did Bloch in *Historian’s Craft*.

Thus analysis was the central work of the historian, and following that, synthesis in the writing of books and articles. This was the classic form of historical method and the analysis had at its heart a shared kindred with the archivist – authenticity. Archivists concerned themselves with the authenticity of the document. Historians concerned themselves with the intentions of the historical actors who created the documents.

In her book, *Dust: The Archive and Cultural History*, Carolyn Steedman reminds us that much of what is written as history is based on research in archival institutions that generally did not exist before the eighteenth century. These archives contain discombobulated remnants of human experience. As opposed to functioning as sources of truth, they are reservoirs of stories that historians use to construct meaning out of the dust and detritus of people’s lives (Ganaway, 2003). This meaning, due to the shards of the past from which it is reconstructed, isn’t truth but a fiction. The typical European state
archives, in particular, tied to a nineteenth century conception of political history, limits scholars as much as it helps them in their efforts to write historical truth. As she writes about archival materials:

> The smallest fragment of its representation... ends up in various kinds of archive and records office... From that, you make history, which is never what was there, once upon a time. (Steedman, 2001, p. 146)

Steedman counsels that academics keep these limitations in mind when researching and writing history. This has in fact, been an important theme in thinking on historical method the past forty years. In *The Order of Things* Foucault suggested we could locate the true origin of modernity in scientists' rejection of form in favor of function as the best means for classifying plants and animals. A corollary of this involved governments gathering documents to codify and categorize their citizens. Furthermore, most scholars in the humanities have read anthropologists Clifford Geertz and James Clifford, or post-colonial thinkers such as Antoinette Burton. They recognize the archive is a creation that masks as much as it reveals, and that regardless of where the author did research his voice shapes the final product (Ganaway, 2003).

Steedman suggests that the archive is interesting and relevant today insofar as it shows us the ways people use the past to define themselves and others. Her definition of an archive is fascinating, she views it simply as a name for the places

> ...in which the past (which does not now exist, but which once did actually happen; which cannot be retrieved, but which may be represented) has deposited some traces and fragments, usually in written form. (Steedman, 2001, p. 69)

Steedman shows that archives are impossible to imagine without a specific vision of history as a mass of material that needs to be ordered
according to certain narratives. This follows modern historical method, along a path starting with Giambattista Vico and moving on to Hayden White. The impetus for this turn in historical study came from outside the field, as the result to attention paid to history, and indeed archives, as a form of language by philosophers, literary theorists and cultural critics. Michel Foucault, Jacques Derrida, Louis Mink, and David Carr; all of whose impact will be more closely investigated during the thesis study, believed that the narration of history was something quite apart from what actually happened. Ranke’s simple belief that the historian was able to reconstruct the past from a thoughtful and thorough study of the archives no longer holds true. Historians study the archives for what is not there: the silences and the absences of the documents (Steedman, 2001, p. 151). This is especially the case for labour historians or gender historians.

But curiously, the “historian’s massive authority as a writer” is still fundamentally derived from archives and the use of archives (Steedman, 2001, p. 145). This is evidenced by the “rhetoric” of history writing such as referencing and footnotes. The authoritative value of history still rests on the use of evidential documents. Steedman concedes this point. Historians such as Susan Grigg, writing in the 1990s, could still effectively claim archives to be the “foundations of historical method” (Grigg, 1991, p. 228). Today there is a reawakening of more traditional views of history as evidenced by the rise of political history at American universities and in the professional community. The debate on the nature of historical writing continues and the thesis will further outline the debate, and the role of archives within the study and writing of history. This will be developed by a study of seminal writings on historical method, those mentioned above and others, and by a study of
articles from such journals as *American Historical Review*, *American Historical Studies*, *Journal of American History*, and *Rethinking History*.

The thesis will also explore the impact of the internet and on-line research on the writing and study of history. Works by Roy Rosenzweig (Rosenzweig, 2001 and Rosenzweig, 2003) and Daniel J. Cohen (Cohen, 2004) will figure predominantly.

### 2.4 Various Archives and History Websites

Various archival and history websites will be examined during the thesis writing process. These will be important in ascertaining exactly the current state of digitization practice. The sites will be identified during the writing of the literature review. They will be identified by the following two criteria. First, each website selected will be maintained by an institution (either an archives, library, museum, or some other agency) and not an individual. Second, each website selected will have been designed to serve a purpose beyond the exhibition of a handful of documents. Each will be an effort to digitally introduce researchers to a larger body of archival documents. To put it another way, each website will be intended as a research website. It is envisioned that about twelve to eighteen websites will be examined in a detailed manner. Three websites are discussed below.

The Northern Research Portal project, at [http://scaa.usask.ca/gallery/northern](http://scaa.usask.ca/gallery/northern), aims at digitizing significant parts of major collections. The site includes interpretive exhibits on specific themes. These are directed at various audiences – in fact it is designed to have different parts of the site and different materials accessible to different groups. The groups are school children, the general public and advanced
researchers. The website includes course materials for classes offered through the University of the Arctic, a cooperative international network of universities, colleges, and other organizations committed to higher education and research in the North. The site is designed with:

1. the ability to browse all materials by subject or by geographical location;
2. a section for on-line resources, including both interpretive exhibits and other resource sites;
3. a database of photographs;
4. finding aids to archival collections, linking where appropriate to the digitized resources;
5. a bibliography of resources, linking to the digitized resources;
6. overall site search capability;
7. a section for teachers; and
8. quick access links for the three identified user groups.

The John Curtin Prime Ministerial Library offers access to primary and secondary source materials relating to Australian Prime Minister John Curtin (at http://john.curtin.edu.au/). Curtin held office from 1941 to 1945 and led the country through the height of World War II. The collection includes photographs, documents and oral histories. Virtually all of the collection is available on-line. Contextual links are effectively maintained as researchers may discover documents using a keyword search and then navigate from the document to others contained in the original file. Other files and other series are easily linked to as are photographs and audio or video files, all with contextual links and full metadata in the catalogue record.

collection the United States Library of Congress digitized over 1,900 glass negatives, over 700 prints for which there are no negatives, and album pages showing all prints in the collection. Prokudin-Gorskii’s unusual triple-frame black-and-white negatives consist of three exposures made through blue, green, and red filters to produce photographs that could be printed or projected in colour. All triple-frame glass negatives in the collection have been digitized.

The site provides: details on the unique photographic process used by Prokudin-Gorskii; a biography of him; historical context on the Russian empire during the period; and a detailed database that links to all images. Links in the database place each photograph in the context of the album from which they originate, as well as other contextual and technical information. The user may search the database directly, navigate through the various series of the collection, or view an exhibition of the photographs. The contextual information details the provenance of the collection and how it came to be deposited at the Library of Congress in Washington DC.

From these examples, several criteria emerge on which one might judge archival and historical websites. These criteria are, first, the completeness of the digitized collection. Projects like the Proudin-Gorskii photographs or the John Curtin Prime Ministerial Library have endeavoured to digitize the complete body of records providing as complete a view of the original documents as possible. Second, the ability to connect the digitized items back to their actual archival collection (physical context). This allows archival researchers to refer back to original documents should they so wish. Third on the list of criteria is the ability to place the materials in their historical context. This gives researchers a clearer understanding of materials
they find on the internet. And finally, the level of instruction provided to users by the website is an important consideration. Clearly explaining how the materials are organized, structured, and presented increases the ability of the user to conduct research effectively.

2.5 Conclusion

This literature review encompasses three main areas – archival theory and writing about digitization, historical method, and an example of current archival websites.

An attempt was made to identify the key sources of archival theory. The review shows that archival literature is rich in writing regarding theory and theoretical development. This will make it feasible to investigate how archival theory has evolved and what the triggers for such evolution have been in the past. There is also a growing body of literature concerning archival digitization. Writers such as Lilly Koltun, David Bearman, and Jennifer Trant are identifying problems posed by digitization. This makes feasible a study of why theory has been ignored in digitization, or, at the least, why there has been a divergence between archival theory and digitization practice.

The same may be said of writing on historical method. A vast body of literature exists on the evolution of historical study and writing. This ranges from classical 19th century primary-source focused methodology to post-modern literary criticism. The development of historical method and archival theory mirror each other in a way. Both modern disciplines date from a 19th century European tradition in which nationality and the state were paramount. Both have placed a heavy emphasis on the evidence of
documents, and recently both have been influenced by the ideas of post-
modernism. An understanding of historical method will be essential to 
understanding important features of effective websites as historical study, of 
one kind or another, is a significant purpose of archives and by extension is a 
significant purpose of use of archival websites.

The examination of archival websites in this literature review shows 
that such an exercise in truly worthwhile and that criteria for website 
selection and evaluation are discernable from such a review.
3 Research Method

3.1 Introduction

Three research methods will be utilized in the conduct of this research project. All three have been extensively used in archival research (Gilliland and McKemmish, 2004, pp. 178, 185-87). Historiography and theory building will be used along with a series of surveys structured according to the Delphi research method.

3.1.1 Historiography and Theory Building

Broadly speaking, historiography is the body of techniques, theories, and principles of historical research that involve critical examination, evaluation, and selection of material from primary and secondary sources (Gilliland & McKemmish, 2004, p. 185). This method will be ideal in collecting and evaluating the body of textual writing on which archival theory is based. These writings stretch back to the early part of the 20th century and a critical analysis of these, one against the other, will clearly show the evolution and acceptance of archival thinking to this day. Historiography is itself the method used to develop a number of writings that will be useful to this study. It was used by Ernst Posner (Posner, 1940) in tracing the development of archival theory since the French Revolution and by Terry Cook (Cook, 1997) in presenting the direction of archival thinking in the 20th century. Other writers, such as Richard Cox (Cox, 2000) and Eric Ketelaar (Ketelaar, 1996; Ketelaar, Horsman, and Thomassen, 2003), have extensively used this method.

As part of this research project, augmented theoretical models for archival digitization may be postulated and explored. A theory building
model will be used, drawing on the existing theories, frameworks, concepts and models explored during the historiography exercise. This will be supplemented by observation, scholarly and professional communication, and data derived from surveys. Theory building also has a strong tradition in archival research design. Writers such as Chris Hurley (Hurley, 1995) and Tom Nesmith (Nesmith, 1999; Nesmith, 2002) have utilized this method extensively.

3.1.2 The Delphi Method

The Delphi method will be used to undertake a survey of archivists and archival users. Although less well known than other research models, the Delphi method nevertheless can be very useful. The Delphi method was chosen as the proposed study group will be too far spread for face-to-face meetings or interviews but not large enough to provide a suitable sample for a survey by other methods.

In a survey conducted according to the Delphi method a panel of experts is selected and a series of questionnaires administered. The compositions of the panel are not disclosed to participants; nor are participants involved in any direct interaction. Communication with the survey administrator is also primarily done in writing. The panel is sent a first-round questionnaire. The results from this first-round, together with an anonymous summary of the comments given by the respondents, are incorporated in a second-round questionnaire and the original questions repeated. Each respondent has the opportunity to review and reconsider his or her own view in the light of other views of the panel. Further rounds may be conducted on the same basis. Anonymity of both the participants and their specific judgments is maintained throughout the whole exercise.
While anonymity and the lack of face to face contact may be a disadvantage of this method – particularly the lack of stimulation and focus that a face-to-face meeting may provide – the systematic bringing together of expert views in several rounds to reach consensus usually provides a stronger basis for decision than simple individual judgments would (Williamson, 2002, pp. 214-17).

Library and information studies use the method regularly. Christopher Buckley (Buckley, 1994, p.159) first realized the value of the Delphi method in assisting with technological change. “With a plethora of emerging new technologies and initiatives to choose from,” Buckley wrote, “Delphi techniques may serve as a worthwhile tool for choosing which to implement.” In a recent archival study, Delphi method was employed to critically evaluate a series of records management toolkits for good corporate and information governance (McLeod, Childs, and Heaford, 2007).

3.2 Research Design

The proposed research design will provide a suitable environment to employ the chosen methodologies.

The literature review will proceed primarily using the historiography method. A review of literature and of practice, through the review or archival and history websites, will be undertaken. The criteria for selection of websites will follow that outlined in section 2.4. As this proceeds, the first stage of questionnaires will be administered.
The Delphi method will be used to undertake a survey of archivists and archival users. A panel of experts will be selected. The criteria for selection will be that these experts must be either professional archivists or extensive research clients of archives who have utilized materials both in person and digitally. The archivists selected for the sample will be involved in creating digital materials or involved in digital reference services. The researchers will be both professional and amateur researchers undertaking both academic and personal research initiatives. These experts will be drawn from Australia, Canada, the United States, and the United Kingdom. They will number forty – twenty archivists and twenty researchers. Archivists will be recruited by contacting archival institutions and asking for interested archivists to participate. Listserv announcements on relevant listservs will be used to recruit archival researchers for the study.

Two stages of survey will be administered. The first will relate to useful elements of a web site, the role of authenticity, and research purposes. It is anticipated that two-rounds of survey will be administered at this stage: an initial round and a second round for participants to refine their responses in light of those provided by others. This is a key component of the Delphi method.

This first stage will be administered early in the thesis research process. The resulting qualitative data and the study of literature and websites will then be analyzed. This critical analysis will be conducted using the theory building model. At the end of this process a theoretical digitization model will be postulated.
A second stage of survey will be undertaken, about one year after the first stage. The same participants will be presented with a draft theoretical digitization model for comment. Again two rounds will be administered and the responses will then be considered in the development of a final theoretical model.

The founders of the Delphi method, Norman Dalkey and Olaf Helmer of the Rand Corporation (Dalkey and Helmer, 1963, cited in Williamson 2002, p. 210) stated that the primary purpose of the Delphi method was to “obtain the most reliable consensus of opinion of a group of experts”. This is exactly the use to which it will be put in this research project. A further aspect of Delphi is its effectiveness (Linstone & Turoff, 1975, p. 3) “in allowing a group of individuals, as a whole to deal with complex problems”. It will be utilized to investigate the use, or lack of use of archival resources by researchers, both as perceived by archival researchers and archivists. This is a complex problem that is difficult to effectively break into a series of less-complex questions.

The questionnaires presented to archivists and to archival users will vary slightly in wording but, in essence, will ask the same questions. The aim of the first stage questionnaire will be to pinpoint what constitutes a useful and effective presentation of digitized archival materials by web site. Rather than asking users what they “want”, questions will be framed so as to ascertain what research they “do” and to what purposes they place the research once it is complete. From this data qualitative information on useful and effective components of web sites, and on the user’s understanding and need for authentic information will be drawn. Similar questions will be asked of archivists to ascertain their understanding of user needs.
3.3 Data Collection and Analysis

Data collected using the survey, and literature and websites identified in the thesis literature review, will be analyzed using the theory building method of critical analysis detailed in section 3.1.1. Drawing on the concepts explored during the historiography exercise, the survey results, literature, and websites will be critically examined against the theoretical precepts of provenance, context, and archival evidence. It is anticipated that all analysis will be of a qualitative nature. The aim of the survey, for example, is to elicit professional and experienced opinion on research using digital archival materials. This opinion, it is hoped, will be thorough and considered.

Questionnaires will be circulated to, and received from, participants electronically. Each questionnaire will be accompanied by an outline of the project which will describe in clear, simple terms, the purposes and procedures of the study. A consent document will be developed. This will be circulated by regular mail.

As part of the Theory Building model, scholars and other professionals will be identified and engaged in communication. These communications will be by email and telephone. Telephone discussions will only be recorded with the advance consent of the participants.

All records will be kept in a confidential manner. Records will not be deidentified as the risk level of information is very low. All responses will be kept and disposed of in a manner consistent with Edith Cowan University practice. The procedures will be submitted to the university’s Human Research Ethics Committee for approval.
4 Conclusion

Archival digitization is an ever growing part of the archival experience. More and more archival material is available on-line each and every day. Archival digitization projects, however, have had a tendency to be selective and to remove records from their context. Long standing archival principles have established that records have value in groups. A record’s ability to serve as evidence depends on its interrelationship to other documents and the removal from this context has negative effects on the researcher.

While researchers are dependent on the soundness of archival records, most archival research is not based on a method that considers provenance, origins, and context carefully. Researchers may not be aware of the loss of context that is rampant with digitized archival materials and the corresponding loss of evidential and research value.

Clearer thinking on archival digitization practice and researcher need is crucial. While there are a handful of archival websites that present archival materials in a contextual way, most do not. This thesis will examine the centrality of archival theory to proper archival practice and will aim to identify the components necessary in building an archival digitization model that is true to archival principles and an effective resource for archival researchers. It is essential that archival professional knowledge be extended to this area of the archival endeavor.
5 References


Cox, R. J. (2000). Closing an era: Historical perspectives on modern archives and


Freeman, E. (1984). In the eye of the beholder: Archives administration from the user’s point of view. *American Archivist, 47*(2), 111-123.


