GETTING PERSONAL: CONFRONTING THE CHALLENGES OF ARCHIVING PERSONAL RECORDS IN THE DIGITAL AGE

by

Jordan Leslie Bass

A Thesis submitted to the Faculty of Graduate Studies of
the University of Manitoba
in partial fulfillment of the requirements of the degree of

MASTER OF ARTS

Department of History (Archival Studies)
Joint Masters Program
University of Manitoba/University of Winnipeg
Winnipeg, Manitoba

TABLE OF CONTENTS

ABSTRACT

Personal digital records are one of the most underrepresented areas of archival theory and practice. Documentary forms created by private persons have long been victim of a poverty of professional attention, and much of the literature on the appraisal and preservation of records has tended to focus on those generated by government and other organizational entities. And strategies developed for the archival management of digital records have similarly placed strong emphasis on business functions or corporate transactions as the primary unit of analysis. This scholastic deficit has severely impaired the ability of the archivist to comprehend and effectively meet the many challenges of archiving personal records in the digital age.

This thesis demonstrates how investigations of the original context of creation and use of records in contemporary personal computing environments are integral to the development of comprehensive strategies for the capture and preservation of personal digital archives. It is within these digital domains that archivists come to see cultures of personal recordkeeping, private appraisal decisions based on unique designations of value, and the complexities of both online and offline personal digital preservation strategies. A keen understanding of how individuals create and preserve their digital records across time and space should be of the utmost importance to archivists for, if nothing else, these pre-custodial activities are the principal sites of archival provenance.

Chapter one discusses past and present responses to both paper-based and electronic personal archives. The discussion begins with the definition of the personal record as essentially non-archival by early leading archival theorists and how these definitions, though first advanced in the early to mid-twentieth century, continue to find

resonance in contemporary archival ideas and institutional mandates. This chapter then illustrates how ideas predicated on the management of electronic government records, and metadata standards developed for formalized electronic recordkeeping systems, are not easily transposed to personal domains. Chapter two takes a critical look at the often oversimplified personal digital archiving environment to expose the many nuances in the context of creation and use of records by individuals in the digital era. Chapter three explores a number of emerging approaches to the professional archiving of personal digital records and reveals how the proper management of these materials requires multiple hardware and software applications, concise acquisition strategies and preservation methodologies, and diligent front-end work to ensure personal digital records cross the threshold of archival repositories. The thesis concludes with a summary of the main arguments and collates the best ideas, approaches, and technologies reviewed throughout to propose a hypothetical strategy for archiving personal digital records in the present.

This thesis argues that significantly more work with records creators earlier in the record creation process must be done when archiving personal digital records because more proactive measures are required to capture and preserve these materials than was previously the case with paper-based or analog documentary forms.

ACKNOWLEDGEMENTS

I would first like to acknowledge and thank Professors Tom Nesmith and Terry Cook. Their passion for archives has been a true inspiration throughout my coursework, internship, conference presentations, and the writing of this thesis. A special thank you to Tom Nesmith, my thesis advisor, for first introducing me to personal archives and for encouraging me to research and write a thesis on this important topic.

I would also like to thank my fellow students from the University of Manitoba Archival Studies Program, the Social Archives Section at Library and Archives Canada, my fellow archivist interns, and now colleagues, from the 'LAC summer of 2009,' as well as my friends and colleagues at the University of Manitoba Archives & Special Collections and the Neil John Maclean Health Sciences Library.

To my family for their love and support, I am eternally grateful.

Finally, I would like to thank my wife, Siobhan Cheekie, for her love, patience, and unfailing support. I could not have written nor successfully defended this thesis without you.

LIST OF FIGURES

List of Copyrighted Material for which Permission was Obtained

Permission Obtained from Seth Shaw, Electronic Records Archivist, University Archives, Duke University Libraries for:

> **Figure 3.1 – Duke Data Accessioner User Interface**. "Screen capture." http://library.duke.edu/uarchives/about/tools/data-accessioner

Permission Obtained from Petar Petrov, Information and Software Engineering Group, Institute of Software Technology and Interactive Systems, Vienna University of Technology for:

> **Figure 3.4 – Home and Office Painless Persistent Long-Term Archiving (HOPPLA) 2.1 User Interface**. "Screen capture." http://www.ifs.tuwien.ac.at/dp/hoppla/

INTRODUCTION

THE VALUE AND PRECARIOUS EXISTENCE OF PERSONAL DIGITAL ARCHIVES

> Today, the will to archive is a powerful impulse in contemporary culture…Today the new information technologies expand our capacity to record everything: to be is to record and record in volume means to classify, index and archive.[1]
>
> People are capturing and storing an ever-increasing amount of digital information about or for themselves, including emails, documents, articles, portfolios of work, digital images, and audio and video recordings. Computer processing, storage, and software tools available to individuals are increasing in power, volume, and ease of use, year on year. Many issues arise from this more informal and increasingly empowered landscape of personal collection,…which will have major future impacts.[2]

Personal archives contain the documentation of individual and collective human experience as witnessed by those who memorialize the activities and events occurring throughout their lives. Though the majority of people may not refer to their collections of letters, diaries, videos, email messages, and photos as archives, these documentary forms are generated by the organic rhythms of everyday life and preserved because of their enduring value which, by many definitions of the word, indeed makes them archival.[3] For individuals, records within their archives serve as a testament to creativity and achievement, are involved in a continued construction and expression of identity, document personal and professional relationships, and facilitate re-encounters with the

---

[1] Mike Featherstone, "Archive," *Theory, Culture and Society*, 23 (2006), p. 595.
[2] Neil Beagrie, "Plenty of Room at the Bottom? Personal Digital Libraries and Collections," *D-Lib Magazine*, 11:6 (June 2005). Available at http://www.dlib.org/dlib/june05/beagrie/06beagrie.html (accessed 21 November 2010).
[3] A standard definition of archives is "Materials created or received by a person, family, or organization, public or private, in the conduct of their affairs and preserved because of the enduring value contained in the information they contain or as evidence of the functions and responsibilities of their creator…" See Richard Pearce-Moses, *Glossary of Archival and Records Terminology*, Society of American Archivists, (2005). Available at http://www.archivists.org/glossary/term_details.asp?DefinitionKey=156 (accessed 25 October 2011).

personal past. For families, personal archives are often a palpable connection between one generation and the next -- a totemic link established and maintained by individual and shared narratives recorded in documentary form. Indeed, the impulses and motivations behind the creation and ongoing curation of a personal archives are just that - personal.

The documentary traces left by individuals are also valued by memory institutions such as archives, libraries, and museums, which seek to acquire personal archives (also referred to as manuscript collections or personal papers) in the interest of cultural heritage, to foster a sense of community, and to develop a rich primary source base for genealogical studies and academic research. In this context, personal archives are defined as the records of prominent authors, photographers and artists, influential religious and political figures, inspiring social activists, or noteworthy leaders in business, medicine, or science. Yet, personal archives are also defined as the documentary forms created and accumulated by individuals whose significance is drawn not as much from their professional accomplishments as it is from what they have recorded from a particularly unique temporal, socioeconomic, or spatial standpoint.[4] In short, the value of personal archives is derived from what they are about as much as it is from who created them and for what purpose.

Possibly the most loyal patron of the archives, the academic historian, has for some time relied on personal archives to provide proximate first-person accounts of events that have occurred in the past in addition to historical narratives that run contrary

[4] Arguably, the significance of "non-famous" donors of archival records may be measured by how those materials resonate within a community. A contemporary example of this is the seven-hundred-forty-one Winnipeg Jets game programs accumulated by Ken Turner, a season ticket holder for the Winnipeg Jets before the franchise left Winnipeg for Phoenix in 1996. They were donated to the University of Manitoba Archives & Special Collections in 2010. The value of this collection increased considerably following the return of the Jets in 2011.

to those put forth by governments and other organizations. Indeed, personal records have figured prominently as primary sources throughout modern historiography, whether in studies of the political and social elite or minority groups and social processes.[5] In addition to academic historians, personal archives such as those generated by literary figures are often sought by researchers looking for early drafts of book or article manuscripts that may reveal new insights into the creative processes of an author. Journeys through the private papers of a literary figure may also yield more intimate details about the life of an author including subtleties such as the interplay between the fictional and the autobiographical voice. Forays into personal archives also elicit intriguing questions such as "How does one choose a single narrative of a life when the archive presents multiple versions? On what basis does one determine which narratives are more 'truthful' than others?"[6]

Although personal archives have often been the province of academic historians and literary scholars, renewed interest in them has come from a more diverse range of users than was previously the case. For example, in researching his own family history, Canadian archivist Robert Fisher notes how the work of genealogists, once dismissed by archivists as amateur, is "growing rapidly and is increasingly well educated" with some "pushing the bounds of traditional family history outward by adopting sophisticated

---

[5] Books by Canadian biographer Charlotte Gray, such as *Canada, A Portrait in Letters*, are examples of using personal archives for historical purposes. Some local examples of the historian's use of personal records may be found in the following works by faculty members of the University of Manitoba, Department of History: Adele Perry, "Is your Garden in England, Sir: James Douglas's Archive and the Politics of Home," *History Workshop Journal* 70 (2010), pp. 67-85.; Esyllt W. Jones and Gerald Friesen eds. *Prairie Metropolis: New Essays on Winnipeg Social History* (Winnipeg: University of Manitoba Press, 2009); and David. S. Churchill, "Transnationalism and Homophile Political Culture in the Postwar Decades," *GLQ: A Journal of Lesbian and Gay Studies* 15:1 (2009), pp. 31-65.

[6] Sally Newman, "Aileen Palmer's Textual Lives," in *The Intimate Archive: Journeys Through Private Papers*, Stephanie Owen Reeder and Tina Mattei, eds. (Canberra: National Library of Australia, 2009), pp. 133-134.

approaches, seeking answers to questions about family relationships, household and daily life, migration and kinship ties, commemoration and memory."[7] This type of lay-historian use of personal archives continues to grow in popularity with the advent of genealogy websites such as *Ancestry.com* and television programs such as *Who Do You Think You Are?* in which celebrities from Canada, the United States, and the United Kingdom search repositories for letters, diaries, photographs and other documentary traces left by their ancestors.

There are, however, uses of personal archives beyond the immediate ones as primary source material for historical research. This point is illustrated by Victor Rosenberg in an exploration of his father's archive. These letters document the experiences of a German Jewish family in Nazi Germany. Rosenberg finds that although the letters are a testimony of the atrocities of the Holocaust and contribute to the broader historical record of events that transpired during that period, his family archives "in an unanticipated and unpredictable fashion, serve as a device for forging contemporary and ongoing familial, interpersonal, and social relationships" and have "laid a new foundation for commemoration and remembrance."[8] In addition, as the research of British archivist Judith Etherton has identified, both the creation and the use of personal and familial archives have been beneficially exploited by health care professionals in their work with patients who are terminally ill or have experienced the traumatic loss of a loved one.[9] Etherton also notes how personal archives find use in cases of parent separation and child adoption where the creation of life story books and memory boxes containing

---

[7] Robert Fisher, 'The Grandmother's Story': Oral Tradition, Family Memory, and a Mysterious Manuscript," *Archivaria* 57 (Spring 2004), p. 127.
[8] Victor Rosenberg, "The power of a family archive," *Archival Science* 11 (2011), pp. 77-93.
[9] Judith Etherton, "The role of archives in the perception of the self," *Journal of the Society of Archivists* 27:2 (2006), pp. 227-246.

photographs, letters, and other personal archival materials are used by social workers to promote a child's sense of identity and belonging throughout often disruptive changes in their life. Thus, while personal archives are often acquired for their potential use in academic research, these materials often have purposes not immediately apparent to nor likely appraised for by archivists.

Ultimately, records created by individuals are acquired and preserved by archives in the interest of formulating a documentary heritage from which elements of social memory, as expressions of shared or collective experience, are informed.[10] As archivist Laura Millar reasons,

> And so, at last, records and archives find their place in the process of memory: as evidence, as memory triggers, as touchstones – acquired, preserved, articulated, and mediated by society in order to contribute to the construction of collective knowledge, identity, and, perhaps, wisdom. Our individual memory gives us our personal past, and our shared past gives us our collective identity.[11]

The value of personal archives - their significance in everyday life, their ongoing benefit to scholarship, their unconventional and ever-evolving utility, and their place in the construction and transmission of collective memory – undoubtedly warrants their inclusion within the repositories of archival institutions. Yet, two decisive factors threaten the survivability of these invaluable resources: the enormous impact of digital technology on how, where, and when documentary forms are created and kept by individuals; and the ability of the archival profession to understand and effectively respond to the unique problems and challenges associated with the appraisal, acquisition, and long-term preservation of personal digital records.

---

[10] Documentary heritage is broadly defined here as surviving documentation (records and publications) of past events that are acquired and preserved in archives. Documentary heritage is not history, but rather materials that are subjected to inquiry, historical and otherwise.

[11] Laura Millar, "Touchstones: Considering the Relationship between Memory and Archives," *Archivaria* 61 (Spring 2006), pp. 125-126.

Over time, the means by which personal archives are constructed have evolved to allow for the creation and gradual proliferation of journal and diary writing, mechanically typed correspondence, photography, self-publishing, as well as the capture of sound and moving images on various types of physical media. With the advent of the personal computer, genres of personal records began to shift from paper-based and analog forms to increasingly electronic and digital formats created in personal computing environments.[12] For instance, conventional written correspondence constructed with pen and paper or mechanical typewriter has by and large been succeeded by word processing programs and text-based computer mediated communications, while the capture of still and moving images is now achieved primarily through digital devices. Beyond this evolutionary transition of conventional personal records from paper to electronic media, however, are documentary forms such as social media posts, email, and blogs, which, as products born purely of the digital age, represent new genres of the archival record.

Personal digital records, like all digital records, rely on supportive hardware and software platforms in order to render their informational content intelligible. The problem is that the hardware and software used in the creation and storage of personal digital records is highly susceptible to rapid obsolescence brought on by commercial forces, and instability in terms of file format corruption and storage media degradation. Unintentional loss of data may also occur as a result of human (accidental deletion) and machine error (hard disk failure), changing terms of service agreements with online service providers (email clients or file sharing platforms), distribution and decentralization of data throughout multiple on and offline storage environments (local hard drives and the

[12] Analog technology involves the process of converting an audio or video signal (transmission of data) into electronic pulses. Digital technology acquires an audio or video signal and converts it into binary format where the data is represented by a series of values. These discrete values are either “1” or “0”.

cloud), or through "benign neglect" when an individual is simply unable to invest the time and effort required for the proper management of their digital records (backup and migration).[13] In short, a number of problems associated with personal digital records stem from the innate ephemerality and instability of digital media.

The second major factor threatening the survival of personal archives in the digital era is the archival profession itself. Archival theory has historically bypassed the realm of personal archives in its privileging of hypotheses and principles devoted to government and corporate records.[14] Furthermore, archival methodologies developed for the capture, management and preservation of digital records have largely been based on formalized record-keeping systems, institutional functions, and the concept of records as evidence of business, government and corporate transactions. Indeed, both paper and digital personal archives have been, as one Canadian archivist aptly notes, "a poor cousin to government archives in the family of archival theory."[15] The downplaying of personal records throughout the broad spectrum of archival theory and practice has severely impaired the development of archival strategies for the appraisal, acquisition and preservation of personal digital records in the twenty-first-century. The tyranny of the digital medium, coupled with the inability of the archival profession to develop comprehensive strategies for management of personal records created in digital form,

---

[13] Catherine C. Marshall, "Rethinking Personal Digital Archiving, Part 1: Four Challenges from the Field," *D-Lib Magazine* (March/April 2008). Available at http://www.dlib.org/dlib/march08/marshall/03marshall-pt1.html (accessed 2 June 2011).

[14] Riva A. Pollard, "The Appraisal of Personal Papers: A Critical Literature Review," *Archivaria* 52 (Fall 2001), pp. 136-150. In this article Pollard points out the ongoing marginalization of personal papers within archival theory as well as the failure of contemporary literature to address the appraisal of personal papers. Although documentary forms are not the primary focus of her article, Pollard defines personal records as "personal papers" which she classifies as letters, diaries, scrapbooks and photograph albums, professional papers, labeled photographs, as well as labeled films, video and audio tapes. There is no mention of personal electronic or digital records in this article.

[15] Robert Fisher, "In Search of a Theory of Private Archives: The Foundational Writings of Jenkinson and Schellenberg Revisited," *Archivaria* 67 (Spring 2009), p. 2.

have given weight to statements from those who believe our documentary heritage has entered a "digital dark age."[16]

Chapter One of this thesis will discuss past and present responses to both paper-based and electronic personal archives. This chapter explores how the personal record was defined as non-archival in foundational archival theory and how these early definitions, first put forth in the early to mid-twentieth century, continue to find resonance in contemporary ideologies and support in the mandates of archival institutions. This chapter reviews and is informed by the work of major contributors to personal archives theory and practice from the 1990s to the present. It will address the questions of why literature on personal archives theory has paled in comparison to that on government and corporate archives, and why after years of pioneering work explicating the uniqueness and value of personal archives, there still remains a considerable scholastic deficit in this area of archival specialization. This chapter will also address the problems archivists have encountered in the application of electronic records management and archiving practices, predicated on government and business functions, to records created by private persons. It will also critically examine the legitimacy and effectiveness of applying rigorous, and yet highly conceptual, authenticity metadata standards to records created not by formalized organizational recordkeeping systems, but rather by individuals operating in informal everyday computing environments.

The second chapter will examine personalized digital archiving environments by reviewing pertinent literature published outside of traditional archival journals and monographs. This chapter seeks to outline the context of creation and use of personal

[16] Terry Kuny, "A Digital Dark Ages? Challenges in the Preservation of Electronic Information," *Sixty-third IFLACouncil and General Conference*, (August 1997). Available at http://archive.ifla.org/IV/ifla63/63kuny1.pdf (accessed 15 January 2011).

digital records, before they are acquired by archival institutions, through the perspective of Personal Information Management (PIM) research to ultimately discover why, how, and where private individuals create and preserve documentary forms in the digital era. This chapter argues that a genuine understanding of the processes of records mediation occurring in the pre-custodial environment of personal digital archives is integral to the discovery and exploitation of their requisite technical and rich provenancial information. To this end, this chapter specifically examines personal digital recordkeeping behaviours and strategies, individual appraisal decisions and designations of value, as well as personal digital preservation practices.

The third chapter of this thesis will explore a number of leading approaches to archiving personal digital records, an area which is only now beginning to be addressed by the archival profession. This chapter analyzes what may only be considered revolutionary approaches to personal digital archives in three key areas: rescuing electronic data from obsolescence; front-end acquisition and preservation methodologies; and embryonic strategies for the future of personal digital archives. Throughout this chapter a number of technological models and emerging software applications are assessed for their real world applicability in the interest of locating practical approaches to the acquisition and preservation of personal digital records by contemporary archival institutions.

This thesis concludes with a summary of the main arguments and proposes a hypothetical personal archiving strategy with the goal of promoting further development in the archiving of personal digital records. As archivists are prime movers in the formulation of documentary heritage;

> They shape to a major degree what society can know about itself. They choose and preserve (or ignore and destroy) recorded evidence of past precedents and societal ideas that are essential to inform the present and guide the future.[17]

This thesis argues that significant upstream effort must be invested in the archiving of personal digital records for the simple reason that more proactive measures are required (earlier in the record creation process) in order to capture and preserve these materials than was previously the case with paper-based and analog documentary forms. A failure to capture and preserve personal digital archives now will assuredly lead to a distorted picture of the past for future generations and severely impair the ability of society to truly know itself.

[17] Terry Cook, "Documenting Society and Institutions: The Influence of Helen Willa Samuels," in *Controlling the Past: Documenting Society and Institutions, Essays in Honor of Helen Willa Samuels*, Terry Cook, ed., (Chicago: Society of American Archivists, 2011), p. 1.

CHAPTER ONE

ARCHIVAL RESPONSES TO PERSONAL RECORDS

> The starting point is theory, which aims to generalize about the nature of archives in order to set the intellectual framework for method and practice. The starting point of theory is to determine the characteristics common to all archives.[1]

The archival profession abounds in both theoretical perspectives and models cultivated to guide everyday practice. Over time, specific formulae have been developed to address problems in the areas of archival appraisal, preservation, and description, among others, in both paper-analog and digital environments. This body of theory has stimulated debate, produced crucial insights, and more importantly, helped generate valuable standards and guidelines which continue to serve as the backdrop to much archival work done today. Unfortunately, as this theoretical base developed a regime of codified principles founded on and directed toward the records created by governmental and corporate entities, it marginalized personal private records and the archivists that work with them. Put differently, much archival method, and the corpus of archival theory from which it is derived, have implicitly ignored glaring discrepancies between records created by governmental and other corporate bodies for purely administrative purposes and those records born of the intimate and otherwise informal everyday needs and desires of private individuals. This chapter will offer a brief history of the divide between personal and institutional archival theory, method, and practice.

This overview begins with an examination of foundational archival theory as laid down in the so-called 1898 "Dutch Manual" on archival administration as well as the ideas of Sir Hilary Jenkinson, and T. R. Schellenberg. These early seminal works defined

[1] Terry Eastwood, "What is Archival Theory and Why is it Important?" *Archivaria* 37 (Spring 1994), p. 129.

as non-archival the records created and kept by private individuals. These early definitions of personal records still permeate the mandates and traditions of conventional archives practice. This chapter then turns first to locating a theory of personal archives, and second, a theory of personal digital archives, by way of a literature review focusing on such topics as archival appraisal and notions of value, post-custodialism, and pre-custodial intervention. This chapter concludes with an examination of how in an effort to protect and sustain the authenticity of the record in the digital age, archivists transfixed by organizational recordkeeping developed a digital preservation framework which, as will be demonstrated by its rigorous metadata requirements, far outstrips the abilities of the everyday individual and the technology they use to create digital records.

Personal Records in Foundational Archival Theory

The records of private individuals have historically been considered inferior to the records of government in archival theory. Early twentieth-century texts focused on the archiving of government and corporate records while the private manuscripts of families and individuals have largely been excluded from consideration in the foundational writings of modern English-speaking archival practice.[2] In the process of excluding private manuscripts from early professional dialogue the progenitors of modern archival theory constructed a lasting discord between the personal and government archival traditions still evident in contemporary archival thought.

The conceptual distinction between personal and government archives finds its antecedents in the period when modern archival principles were beginning to be prescribed in a number of foundational treatises. These early works defined archives

[2] Personal records are also referred to by archivists as private manuscripts, personal papers, personal fonds, historical manuscripts, and manuscript collections.

exclusively within a structured governmental context where records reflected, and were inextricably linked with, the administrative or business functions that created them. For example, in the 1898 *Manual for the Arrangement and Description of Archives*, Dutch archivists Samuel Muller, Johan Feith and Robert Fruin denied archival status to the records of individuals on the grounds that these materials were not the natural result of the functions and activities of an administrative body, but rather were "gathered together in the strangest manner" and better suited for library custody.[3] A landmark in archival literature, the *Manual* was decisive in the development of Western archival theory for its advancement of the concepts of provenance and original order but also for its characterization of personal records as non-archival due to their un-organic, and largely inferior, means of creation.[4]

Building on the core concepts of the Dutch Manual, subsequent influential works of archival theory perpetuated the definition of archives in purely administrative terms while maintaining the subordinate status of private manuscripts. In *A Manual of Archive Administration* published in 1922, British archivist Sir Hilary Jenkinson formed a definition of archives predicated on administrative transactions, the interrelated concepts of authenticity and impartiality, and emphasized the importance of these elements in the cultivation and conservation of the objective evidence of archives.[5] Measured against this

---

[3] S. Muller, J.A. Feith, and R. Fruin, *Manual for the Arrangement and Description of Archives* (1898), 2d ed. (1940) Arthur H. Leavitt, trans. (New York, reissued 1968) pp. 19-22 & 152-155. An archival collection was defined as a living organism, the natural product of functions and activities which formed an *organic whole.* Arbitrarily accumulated historical manuscripts lacked the *organic bond* of archival collections which evolved in accordance with "fixed rules" – as the functions of the administrative body changed, so too did the nature of the archival collection created by that body.

[4] Ibid., pp. 33-35 & 52-59. The principle of provenance is described in *Rule 8* of the *Manual* and the concept of original order *Rule 16*.

[5] Hilary Jenkinson, *A Manual of Archival Administration*, new and revised edition, (London: Percy, Lund, Humphries & Co Ltd, 1937), pp. 11-12. In his definition of archives Jenkinson reaffirmed the theoretical orientation of the Dutch Manual in stating: "A document which maybe said to belong to the class of

definition, private manuscripts imbued with expressions of personal opinion and created in the interests of posterity outside of official business capacities, could not purport to hold the same archival or evidential quality found in those records born of administrative functions.[6] An essential element of private manuscript collections, and perhaps the most troubling aspect of archival practice for Jenkinson, was the occurrence of subjective judgment involved in the selection or appraisal of records, for in his view, these inherently biased acts fundamentally undermined the evidential quality of archives.[7] Archives were kept, not acquired, and the collection of private manuscripts invariably involved selective acquisition.[8]

In the 1940s 1950s, archivist Theodore R. Schellenberg of the American National Archives developed new approaches to the management of records in the National Archives. It had assumed responsibility for a colossal amount of modern government records created since the birth of the republic and especially in the New Deal and World War II eras.[9] In terminology reminiscent of Jenkinson, Schellenberg formulated a precise definition of archives wherein the haphazard creation and

---

Archives is one which *was drawn up or used in the course of an administrative or executive transaction (whether public or private) of which itself formed a part; and subsequently preserved in their own custody for their own information by the person or persons responsible for that transaction and their legitimate successors*."

[6] Ibid., pp. 4-15, & pp. 40-44. Responsible custodianship refers to archives being preserved within the custody of the original creator and its legitimate successors. The responsible and unbroken custody of archives ensured that records were free from forgery and falsification and thus upheld their authenticity. The impartiality or truthfulness of an archives is derived from their being part and parcel of administrative or executive transactions. Jenkinson acknowledged archives would be used for different purposes than that for which they were created, he asserted the records were "impartial" in that these considerations of future use were not present at the time of the records creation.

[7] Ibid., pp. 3-4.

[8] Hilary Jenkinson, "The English Archivist: A New Profession," (l947), in Roger H. Ellis and Peter Walne, eds. *Selected Writings of Sir Hilary Jenkinson* (Gloucester. 1980), p. 238. Quoted in Richard Stapleton, "Jenkinson and Schellenberg: A Comparison," *Archivaria* 17 (Winter 1983-84), p. 77. In a very explicit definition on the growth and accumulation of archives, Jenkinson noted "Archives are not collected: I wish the word 'Collection' could be banished from the Archivist's vocabulary if only to establish that important fact."

[9] Robert Fisher, "In Search of a Theory of Private Archives: The Foundational Writings of Jenkinson and Schellenberg Revisited," *Archivaria* 67 (Spring 2009), p. 12.

unsystematic collection of historical manuscripts stood in sharp contrast to government archives that grew out of regular administrative "functional activity."[10] To meet the challenges posed by the volume of materials generated by modern government, Schellenberg developed a value-based system to assist in the appraisal of archival records. Within this system, records were tested for *evidential value*, found in those records documenting the functions of an administrative body, and *informational value* observed in those records containing information on the people, places, and subjects with which the administrative body interacted. [11] Here, the assignment of evidential value is determined by the relationship between the record and the government activity creating it; whereas private manuscripts "have a meaning of their own without relation to their source."[12] The informational value of records, on the other hand, is determined solely on the basis of its content where the source of creation ceases to be a deciding factor.[13] While the appraisal criteria of Schellenberg acknowledged the possible extension of informational value to private manuscripts, thereby conferring a degree of archival character, it is nevertheless a theoretical construction firmly rooted in government activity which finds no parallel in the lives of private individuals.[14]

The concepts of authenticity, impartiality, evidential and informational value, and records as products of administrative functions are firmly entrenched in the corpus of

---

[10] T.R. Schellenberg, *Modern Archives: Principles and Techniques* (Chicago: University of Chicago Press, 1956) pp. 13-18. Properly constituted archives were "produced in the course of purposive and organized activity, if they were created in the process of accomplishing some definite administrative, legal, business or other social end" and "prepared for reasons other than those for which they were created or accumulated."

[11] Ibid., pp. 139-141. Records used in the ongoing administrative activities of their creators held primary value. Evidential and informational values form a broader concept of secondary value which was found in all records used for purposes beyond operational activities.

[12] Ibid., p. 141.

[13] Ibid., p. 148. Informational value is also referred to as research or content value.

[14] Fisher, p. 22. Fisher notes that Schellenberg "consistently emphasized the importance of evidential value in the appraisal of government records." The informational value of private manuscripts is a secondary consideration.

modern archival theory and as such have profoundly influenced archivists working in both the government and private dimensions of the profession. Yet, as Canadian archivist Robert Fisher asserts, in their discussions of what characteristics distinguished government and personal archives, the ideas of Jenkinson and Schellenberg "went some distance toward laying a theoretical foundation for understanding private archives, even if it was often expressed in the negative or in terms of absence."[15] In concluding his reexamination of Jenkinsonian and Schellenbergian ideals, Fisher contends that three principles of explanation are used by the founders of English archival thought to define personal archives. First, personal archives are the product of haphazard, perfunctory or spontaneous activities of individuals or informal groups of individuals inhabiting unofficial and unstructured environments. Second, in terms of custody, personal archives are maintained in circumstances where individuals appraise, edit, retain, and destroy documents which are then consequently preserved in individual, familial, or otherwise informal custodianship. And third, personal archives are collected by institutions solely for their research value and potential use by clientele where historical and cultural interests take precedence over evidence of business transactions or government functions. Many archivists today continue to treat these early archival concepts and characteristics as axioms of the discipline and regard evidential and informational value as "standard and present in all archival documentation."[16] As a corollary, this divisive archival ideology has manifested itself in the practices of institutions where the responsibility of archiving is bifurcated in public and private sector acquisition mandates.

---

[15] Ibid., p. 18.
[16] Catherine Hobbs, "Reenvisioning the Personal: Reframing Traces of Individual Life," in *Currents of Archival Thinking*, Terry Eastwood and Heather MacNeil, eds. (Santa Barbara: Libraries Unlimited, 2010), p. 215.

Institutional and Ideological Divisions

In many countries, institutional archival practice has often been based on two competing perspectives: one that defines the archival mission primarily as the preservation of public, corporate, or government records; and one that defines the archival mission primarily as the preservation of personal records. Within these two positions the term "institutional archives" is often synonymous with transactional evidence, accountability, and authenticity, while the term "personal archives" is usually associated with notions of history, heritage, culture and memory. These understandings of archives have found prominent expression in the United States, Great Britain, and Australia for example, where, at the national level particularly, archival responsibilities for government and personal records have been divided between their national archives and library repositories.[17] However, the Canadian "total archives" tradition embodies a more holistic approach which recognizes both personal and institutional records as being of the same archival value and stands as a model for more pluralistic and inclusive conceptions of archival practice.

Total archives is a particular approach to archival management where centralized and publicly funded archival institutions acquire and preserve both the records of

[17] Luke J. Gilliland-Swetland, "The Provenance of a Profession: The Permanence of the Public Archives and Historical Manuscripts Traditions in American Archival Theory," *The American Archivist* 54 (Spring 1991) pp. 160-175. Rebecca Hirsch, "The Permanence of Provenance: The 'Two Traditions' and the American Archival Profession," *Journal of Archival Organization* 8:1 (2010), pp. 54-72. The archiving of documentary heritage in the United States originated not with the establishment of its National Archives in 1934, but rather in its historical manuscripts tradition where historical societies collected documents and other antiquities in the interests of posterity, a tradition dating back to the early nineteenth-century. The American historical manuscripts tradition was later institutionalized at the Library of Congress, for example, where personal manuscripts of national significance are housed. Adrian Cunningham, "Chapter 2: Archival Institutions," in *Archives: Recordkeeping in Society*, Sue McKemmish, Michael Piggott, Barbara Reed and Frank Upward, eds. (New South Wales: Centre for Information Studies, Charles Sturt University, 2005), pp. 21-50 and 33-44.

government and individuals irrespective of their medium.[18] From its inception, the archival mission within Canada has focused on the acquisition of collections of personal papers and other privately generated materials for historic and cultural purposes where the preservation of records as evidence of government activity received little, if any, consideration until the 1950s.[19] Furthermore, the perception that government records warranted archival preservation was based not on a recognition of the need for their proper management, but rather on their historical importance and research value.[20] By the mid-twentieth century the concept of archives in the Canadian context had evolved from a vaguer eclectic commitment to historical 'materials' to confer equal importance on both private and public archives in an effort to document all facets of Canadian life, making it fundamentally different from American, English and Australian definitions of archives. By the time the concept was formally articulated by Dominion Archivist Wilfred Smith in 1972, total archives had grown to reflect the diversification and diffusion of archival labour in national, provincial, municipal, university, business and local repositories, a collective responsibility for the preservation of public and private documentary heritage.[21] In essence, the physical separation of public and private records seen in American, Australian and British repositories may be viewed as the institutionalization or official sanction of the exclusionary practices first put forth in foundational archival

[18] Laura Millar, "Discharging our Debt: The Evolution of the Total Archives Concept in English Canada," *Archivaria* 46 (Fall 2001), pp. 103-146. Publicly funded archival institutions include national and provincial-level archives.
[19] Ian E. Wilson, "A Noble Dream: The Origins of the Public Archives of Canada," *Archivaria* 15 (Winter 1982-83), pp. 16-35. Terry Cook, "An Archival Revolution: W. Kaye Lamb and the Transformation of the Archival Profession," *Archivaria* 60 (Fall 2005), pp. 185-234. The balanced approach to historical records acquisition and government records programs was established during the tenure of Dominion Archivist Dr. W. Kaye Lamb (1948-1968) who reconciled the cultural mandate of the Public Archives of Canada (PAC) with the management of vast quantities of government records generated during the Second World War.
[20] Millar, p. 113.
[21] Ibid., pp. 117-139.

theory. Within this context, however, the concept of total archives emerges as a uniquely Canadian contribution to personal archives theory and practice in its concordant, rather than divisive, definition of what records constitute archives.

Although standing in sharp contrast to other national archival traditions, the Canadian archival experience has not been immune to protracted ideological tensions brought on by the separation of public and private records in conflicting interpretations of the archival mission. Australian archivist Adrian Cunningham aptly describes this ideological tension as the "flinty relationship between archivists who collect the private records of individuals and the rest of the archival profession…" and acknowledges how definitions of the archival record in purely transactional terms are "symptomatic of the corporate myopia afflicting many of today's archival theoreticians."[22] Cunningham argues that a focus on accountability and administrative efficiency narrowly defines archiving as the preservation of organizational records as evidence and diminishes the capacity of archives to function as a cultural entity. In ensuring the preservation of evidence for the purpose of organizational accountability, cultural and historical considerations are dismissed from archival purview where personal records are cast "beyond the pale". Similarly, American archivist Mark A. Greene identifies this professional polarization between what he calls the archival paradigm (encompassing both personal and institutional records) and the recordkeeping paradigm, "which posits that archives are records, *but* that records are solely evidence of transactions, that they are kept primarily (some argue solely) for purposes of administration, law, and

[22] Adrian Cunningham, "Beyond the Pale? The 'flinty' relationship between archivists who collect the private records of individuals and the rest of the archival profession," *Archives & Manuscripts* 24 (May 1996), pp. 20-26.

accountability."[23] Within this archival dialectic, proponents of the archival paradigm acknowledge personal records as essential in the construction and perpetuation of history, heritage, culture and memory whereas the recordkeeping polarity sees the acquisition and preservation of privately generated materials as less socially relevant and quite secondary to the maintenance of evidence and accountability.[24]

The institutional and ideological division of archives has given rise to questions regarding the priority of acquiring personal archives within the construction of documentary heritage. For example, in 1994 Canadian historian Robert A.J. McDonald asserted that the ascendancy of the recordkeeping paradigm had come at the expense of personal archives and suggested that a "tide of diminishing commitment" on behalf of publicly-funded institutions to collect personal records will have a negative impact on collective memory. The Canadian archival profession, as viewed by McDonald, had come to place greater emphasis on records management and the preservation of the institutional record while in the process losing sight of its broader historical and cultural mandate.[25] The reduced priority of personal archives acquisition and a depreciation of values associated with personal records are in part due to a broader professional devaluation of the cultural, heritage, and historical role of archives.

---

[23] Mark A. Greene, "The Power of Meaning: The Archival Mission in the Postmodern Age," *The American Archivist* 65 (Spring/Summer 2002), pp. 42-45. Emphasis in original.

[24] Ibid., pp. 47-51. Greene says that the origins of the recordkeeping paradigm are firmly rooted in the ideas of Jenkinson, most notably in his assurance that true archives are authentic, reliable, impartial records of transactions.

[25] Robert A.J. McDonald, "Who is Preserving Private Records?", "Acquiring and Preserving Private Records: Cultural versus Administrative Perspectives," and Christopher L. Hives "'Thinking Globally, Acting Locally'," *Archivaria* 38 (Fall 1994), pp. 155–163. In a series of three articles, McDonald and Hives debate the priorities of publicly funded archival institutions when it comes to the acquisition and preservation of personal records. While McDonald saw a diminishing commitment on behalf of institutions to acquire private materials, Hives firmly believed that individuals and local organizations should take more responsibility for the management and preservation of their own records.

To some, the downplaying or "abdication" of the cultural role of archives continues to be equated with a loss of information vital to holistic documentary heritage and an understanding of the past as viewed by non-governmental entities.[26] For example, in 2011 the Canadian Association of University Teachers (CAUT) criticized the private records acquisitions practices of Library and Archives Canada (LAC) stating "[t]he fear is that LAC is being reduced to collecting government papers and not much else."[27] Though CAUT mistakenly depicts LAC as being *the* central repository for the totality of Canadian documentary heritage and effectively ignores the true collaborative inter-institutional nature of the Canadian total archives tradition, a steady decline in acquisitions from private sources with an increase in the number of government acquisitions at LAC from 2007 to 2010 does question the commitment of the institution to the preservation of personal archives.[28]

Given the duality of practice, the duplicity of discourse, and the subordinate status of personal records in foundational archival texts, personal archives theory has developed along very different lines from the archives of government. Compared with the wealth of literature on topics associated with the records of government, a deficit in scholarship on personal archives becomes a glaring reality. However, there remain a number of pioneers

---

[26] Shirley Spragge, "The Abdication Crisis: Are Archivists Giving Up Their Cultural Responsibility?" *Archivaria* 40 (Fall 1995), pp. 173-181.

[27] Canadian Association of University Teachers (CAUT), "Open Letter to Daniel Caron, Librarian and Archivist of Canada," (June 2011). Available at http://www.caut.ca/uploads/CAUT_to_Caron_LACv2.pdf (accessed 17 October 2011).

[28] CAUT, "Open Letter to Daniel Caron". CAUT is misinformed in its understanding of both LAC's mandate and the total archives tradition. CAUT sees the decentralization of archival acquisition, preservation, and access as something unique to the 2009 LAC Modernization Project, and yet LAC has never been the "central collector of national resources" as the acquisition of Canadian documentary heritage is far beyond the capabilities of one sole archival organization, even one as large as LAC. Daniel Caron is the Librarian and Archivist of Canada and thus head of LAC. Statistics on the decline in private records acquisitions by LAC are available at Government of Canada, "Libraries and Archives Canada, 2010-2011, Report on Plans and Priorities," See "Section I – Department Overview, Responsibilities, Acquisition trends." Available at http://www.collectionscanada.gc.ca/012/012-205-e.html (accessed 17 October 2011).

in this area of specialization who have succeeded in the creation of definitions and conceptual standpoints unique to the records of individuals that serves as the foundation for future developments in personal archives theory and practice. This chapter now turns to an examination of the beginnings of a theory of personal archives.

Locating Theory on Personal Archives

Archival responses to personal records are most evident in the discourse on archival appraisal, which has focused on two core issues: the matter of creating a methodology to assist in the identification of those individuals within society whose records warrant archival preservation; and second, the matter of producing adequate criteria for the selection of documentary forms within personal archives. Traditionally, archivists have approached the appraisal of personal records in the same fashion as they do government records – that is – through the evaluation of evidential value. For example, American archival educator and theorist Richard Cox asserts:

> An individual maintains records for generally the same reasons as does an organization – to meet the needs of accountability, evidence, and corporate memory. Personal records are created out of the same needs to capture transactions, document activities, serve legal and administrative functions, and provide a basis for memory. We maintain records to create our own evidence.[29]

In addition to stressing the evidential quality of personal records, Cox also maintains the "vast majority of personal and family papers are records with the same organic, orderly nature deriving from functions and activities as institutional records," which in the realm of the personal are defined as "the impulses driving individuals and families to create, maintain, and use their own records."[30] In short, Cox does not view the recordkeeping of individuals as being different from the recordkeeping of governments, which leads to his

[29] Richard J. Cox, "The Record in the Manuscript Collection," *Archives & Manuscripts* 24:1 (May 1996), p. 52.
[30] Ibid., pp. 51-52.

assertion that both public and private records may be appraised for the same evidential value.

The evidential quality of personal records is also explored by Sue McKemmish in her influential article titled "Evidence of me…" which reinforces a definition of personal archives as much the same as the corporate variety:[31]

> Recordkeeping is a kind of witnessing. On a personal level it is a way of evidencing and memorializing our lives – our existence, our activities and experiences… Archivists can analyse what is happening in personal recordkeeping in much the same way as they analyse corporate recordkeeping. Just as they can identify significant business functions and activities and specify what records are captured as evidence of those activities, so they can analyse socially assigned roles and related activities.[32]

McKemmish further explains that private records creation is driven by a dual impulse to construct a sense of self or 'place' in the world and to "function effectively" in socially assigned roles, where records provide faceted information about individual life but "evidence first and foremost" the interaction between the individual and society.[33] McKemmish also suggests that socially conditioned recordkeeping behaviours of individuals yield correlative documentary forms, which in turn communicate different societal roles and aspects of life. McKemmish ultimately asserts that personal records should be incorporated into the "collective archives" to ensure "evidence of me" is carried beyond the boundaries of individual life and transformed into "evidence of us." Here then, the study of personal recordkeeping behavior and the motivations behind records creation are integral to the appraisal of personal records for their unique, but nevertheless evidential, value. In summary, McKemmish views personal archives in terms of functionality and evidence or, in other words, "how systematically we go about

---

[31] Sue McKemmish, "Evidence of me…," *Archives & Manuscripts* 24:1 (May 1996), pp. 28-45.
[32] Ibid., pp. 29-30.
[33] Ibid., pp. 30-38.

the business of creating our records as documents, capturing them as evidence…, and keeping and discarding them over time."[34]

Richard Cox and Sue McKemmish advocate for a particular approach to personal archives where government recordkeeping theory is unproblematically transposed to the realm of the personal based on the universality of recordkeeping activity in society and the omnipresence of evidential value in archives. It is an approach which conflates personal recordkeeping cultures with government recordkeeping models in an effort to structure and impose order on an area of archives otherwise limited in actionable theory and practical application. This approach, however, is challenged by South African archivist Verne Harris who questions the oversimplification of personal recordkeeping and archives in terms of functionality and evidence. Harris argues that personal records creation "is contaminated by the human instinct to tell story and to create identity" where occurrences of imagination, fabrication, forgetting, and healing – all inherent characteristics of personal records - "cannot submit to an economy of proof" and may only be "squeezed into the claustrophobic space of recordkeeping functionality at a price."[35] Personal archives, according to Harris, are more about storytelling, memory or "remembrancing", than they are the systematic capturing of transactional evidence. For Harris, the realm of personal archives "is one fraught with complexity"; a "wilderness" to be acknowledged, respected, and conserved rather than controlled by standards and

[34] Ibid., p. 29.
[35] Verne Harris, "On the Back of a Tiger: Deconstructive Possibilities in 'Evidence of Me,' Available at http://www.mybestdocs.com/ (accessed 15 May 2011). Also published in *Archives and Manuscripts* 29 (May 2001), pp. 8-22.

systems predicated on notions of organizational functionality and structured recordkeeping.[36]

Echoing the sentiments of Verne Harris, Canadian literary archivist Catherine Hobbs attempts to redirect the discourse on personal archives beyond socially conditioned recordkeeping behaviors and concepts of evidence, toward discussions of the interiority of personal records and a strategy for appraising value unique to personal archives not always evident in the records of government and corporations. Although Hobbs does acknowledge personal records convey evidence of the public and transactional activities of individuals, she asserts these materials:

> eclipse both evidential and informational value by their narrative value: they are in many senses creations of the self and participate in a process of storytelling and de-facto autobiography – of the self presenting or representing the self...[personal archives] is a site where personality and the events of life interact in documentary form.[37]

Hobbs argues that narrativity, idiosyncracy, intimacy, spirituality, aesthetics, as well as expression of character and emotion are values unique to documentary forms created by individuals and as such warrant different and more nuanced appraisal criteria for personal archives. Furthermore, as personal records are created to serve the needs or personality of the individual, not government, they indeed provide the most "prevalent source of commentary on daily and personal life and relationships, almost by their very nature."[38] Hobbs rejects the idea of applying government-based recordkeeping practices to personal archives as advocated by Cox and McKemmish, and calls for a departure from archival

---

[36] Ibid.
[37] Catherine Hobbs, "The Character of Personal Archives: Reflections on the Value of Records of Individuals," *Archivaria* 52 (Fall 2001), pp. 127, 131.
[38] Ibid., p. 127.

approaches hinging on the transactionality of the record and the concept of evidence.[39] Like Harris, Hobbs believes in conserving the "wildness" of personal archives rather than taming it.

While the dialogue on personal archives has strived to articulate the distinctive values associated with the records of individuals, it has also engaged the issue of identifying those within society whose lives and corresponding records warrant archival acquisition and preservation. Traditionally, appraisal strategies have been "so random, so fragmented, so uncoordinated, and even so often accidental" in the private sector as to suggest the absence of any strategy at all.[40] Seeking to discover an appraisal theory of personal archives, American archivist Riva Pollard critically examines modern appraisal models employed for government and corporate records and investigates their relevance and applicability to personal records.

Drawing on the influential work of German archival appraisal theorist Hans Booms, who argued the appraisal of archives should be based on scales of value determined by the society and era in which the records were created, Pollard suggests archivists should work to better understand the functions of records creators and the relationships between those functions and the records created. [41] To further illustrate this point, Pollard refers to macro-appraisal methodology, which holds that "the social context of the record's creation and contemporary use (not its anticipated use) establishes

---

[39] Hobbs, "Reenvisioning the Personal," p. 221.
[40] F. Gerald Ham, "The Archival Edge," *The American Archivist* 38 (January 1975), p. 5.
[41] Hans Booms, "Überlieferungsbildung: Keeping Archives as a Social and Political Activity," *Archivaria* 33 (Winter 1991-92), pp. 31-33, and "Society and the Formation of a Documentary Heritage: Issues in the Appraisal of Archival Sources," *Archivaria* 24 (Summer 1987), pp. 76 & 104. "We believe that only the society from which the material originated and for whose sake it is to be preserved can provide archivists with the necessary tools to assess the conceptions by which they bring past to present…archivists must not follow the value concepts of their own time period, but rather, those of the time from which the material originated…the question of value ascribed by those contemporary to the material should become the most fundamental of every archival endeavour to form the documentary heritage."

its relative value."[42] By focusing on "why records were created rather than what they contain, how they were created and utilized by their original users rather than how they might be used in the future." archivists may come to appreciate the functions, roles, motivations, and behaviours behind personal records creation and leverage this newly found appreciation in their appraisal decisions.[43] But as macro-appraisal situates the personal records creator in such a way as to give priority to records documenting first and foremost their interactions with government, Pollard credits documentation strategy, an appraisal methodology pioneered by American archivist Helen Samuels, as a prospective approach to document the more salient activities and events occurring beyond the parameters of citizen-state interactions.[44]

> Documentation strategy is defined by Samuels as:
>
> > a plan formulated to assure the documentation of an ongoing issue, activity, or geographic area…The strategy is ordinarily designed, promoted, and in part implemented by an ongoing mechanism involving records creators, administrators (including archivists), and users. The documentation strategy is carried out through the mutual efforts of many institutions and individuals influencing both the creation of the records and the archival retention of a portion of them.[45]

In order to identify those individuals within society whose records warrant archival acquisition, the documentation strategy looks first at the themes, issues, and activities within society to be documented and the information required to do so. It then turns to an examination of all available forms of documentation, irrespective of their medium and means of creation, and assesses their ability to provide the desired information on a given

---

[42] Terry Cook, "Macro-appraisal and functional analysis: documenting governance rather than government," *Journal of the Society of Archivists* 25, no. 1 (2004), p. 8.
[43] Terry Cook, "Mind Over Matter: Towards a New Theory of Archival Appraisal" in *The Archival Imagination: Essays in Honour of Hugh A. Taylor*, Barbara L. Craig, ed., (Ottawa: Association of Canadian Archivists, 1992), p. 111.
[44] Macro-appraisal is first and foremost a methodology concerned with appraising records of government or other institutions.
[45] Helen Samuels, "Who Controls the Past?" *The American Archivist* 49 (Spring 1986), p. 115.

theme, geographical area, or societal issue.[46] Following in-depth analysis and identification of the activities necessary to document a specific societal topic or theme, the documentation strategy aims to locate those key individuals operating within those identified activities -- at which point the acquisition of personal records occurs. At the heart of the documentation strategy ideal is the belief that "No longer could archivists remain content with their existing collections or with only collections that ended up on their doorsteps; their roles had expanded to include active documentation of society."[47] In short, the documentation strategy depends on rigorous ongoing research that precedes archival records acquisition, which in the realm of the personal translates into a thorough investigation of the motives, activities, and recordkeeping cultures involved in the creation of private records.

In conclusion, archival theory about personal records is divided into two schools of thought: one that believes the concepts of organizational recordkeeping are fully applicable to personal archives, and one that challenges the portability of concepts conceived in organizational settings to the realm of the personal. The first school of thought maintains that the functions and activities of private individuals are indeed equivalent to the transactional business functions of organizational entities. This first school endeavors to bind distinct documentary forms with specific recordkeeping behaviours in an effort to assert the primacy of evidence over more intrinsic values

---

[46] Ibid., pp. 116-124. The documentation strategy further suggests that archival institutions engage in cooperative development of acquisition goals in order to create a representative record of areas identified in their respective acquisition mandates.

[47] Elizabeth Snowden Johnson, "Our Archives, Our Selves: Documentation Strategy and the Re-Appraisal of Professional Identity," *The American Archivist* 71 (Spring/Summer 2008), p. 195. See also, Larry J. Hackman and Joan Warnow-Blewett, "The Documentation Strategy Process: A Model and a Case Study," *The American Archivist* 50 (Winter 1987), pp. 12-47, and Jennifer A. Marshall, "Documentation Strategies in the Twenty-First Century?: Rethinking Institutional Priorities and Professional Limitations," *Archival Issues* 23:1 (1998), pp. 59-74.

associated with personal records. The second school of thought contends that personal archives document the public persona and outward interactions of individuals with others but also a myriad of complex emotions internal to human experience. Personal archives in this second school of thought are envisioned in such a way as to reject any imposition of structural recordkeeping models so as to invite the consideration of the array of human behaviour and motivations involved in the creation of personal records.

Locating Theory on Electronic Personal Archives

Discourse on solutions to problems associated with archives in the information age has brokered many approaches to electronic recordkeeping and the preservation of electronic archives. However, the strategies developed for electronic archives are directed toward government or large corporate institutions and tend to treat all electronic records as products of functions and transactions, giving little if any consideration to the unique characteristics of electronic materials generated in private environments. For example, the task of one of the most influential electronic records projects of the 1990s, the University of Pittsburgh Project, was to identify the functional requirements for electronic recordkeeping in governmental and other organizations, with electronic records explicitly defined as the evidential product of business transactions.[48] Through its analyses of variables in business functions, organizational structures, technological environments, and risk management principles, the findings of the Pittsburgh Project provide a policy-laden schematic to support corporate accountability, but undeniably

[48] David Bearman, *Electronic Evidence: Strategies for Managing Records in Contemporary Organizations*, (Pittsburgh: Archives & Museum Informatics, 1994), pp. 15-17. “In discussion of the guidelines and in work which has followed, the author has depended on a definition of records as business transactions. That is records are transactions which have a significance in business terms (rather than in computing terms) because they constitute evidence of a business event, such as making a sale or qualifying a client to receive a benefit." David Bearman and Richard Cox were the primary investigators of the Pittsburgh Project.

ignore the record creating activities of individuals. For example, it is highly improbable that a private records creator in everyday society would strictly adhere to the first requirement of the Pittsburgh model which is compliance with the laws, regulations and statements of best practice that preside over his or her environment.[49] As Terry Cook asks "what are we to do with the novelist or diarist, the artist or composer, the photographer or preacher…or a hundred others who commit their thoughts, their perceptions, their 'records' of activities to electronic media, and yet do not communicate many of them as 'business acceptable communications' or transactions?"[50]

Given the already inferior status of personal records within the larger theoretical framework of archives discussed in the first section of this chapter; literature on the archival management of electronic records is expectedly derived from an institutional records perspective. Accordingly, the second section of this chapter examines the response to personal electronic records within the larger rubric of electronic archives beginning with an examination of post-custodial theory of archives and the implications this concept holds for personal electronically generated documentary forms.

Post-Custodial Theory and Archival Intervention

Acknowledging the impact of electronic technology on records creation and the changing archival landscape in the 1990s, David Bearman states:

> Archivists need to make a radical break with their past practices to assume responsibility to regulate, to control, the maintenance of the archival record. Unfortunately, archivists have yet to show the willingness to assume this authority or the ability to move beyond a custodial role. Of course archivists will always

[49] Ibid., pp. 293-304. The Pittsburgh Project developed thirteen requirements for electronic recordkeeping systems which may be divided into three broad categories: compliant organizations; accountable recordkeeping systems; and requirements pertaining to the record sub-divided into captured, maintained, and usable. An in-depth discussion of these the thirteen functional requirements is beyond the scope of this chapter.

[50] Terry Cook, "The Impact of David Bearman on Modern Archival Thinking: An Essay of Personal Reflection and Critique," *Archives and Museum Informatics* 11:1 (1997), p. 29.

> end up caring for some records – some they already have in custody and others that no one else can take, but the archival repository is as a strategy for the electronic age what the walled city became with the invention of gunpowder, an indefensible bastion and a liability.[51]

Post-custodial theory of archives, or postcustodialism, is a response to the impact of twentieth-century technologies on means of communication and the increased volume of recorded information. This response signals a reorientation of archives from the custodial era where a focus on the management of the physical document at the end of the records life cycle and the appraisal of its informational content is replaced by an emphasis on the proactive documentation of provenancial and contextual elements closer to the point of records creatorship.[52] Moreover, proponents of postcustodial theory argue that organizational, professional, economic and societal complications involved in the preservation of electronic records are insurmountable outside of their original creation environment. Archival custody of electronic records in centralized repositories is therefore replaced by the decentralization and distribution of custody and access within the originating institution where archivists provide management oversight in the areas of appraisal, documentation and access.[53] Post-custodial theory and its applicability to the

---

[51] David Bearman, "An Indefensible Bastion: Archives as a Repository in the Electronic Age," in *Archives and Museum Informatics Technical Report 13 (1991)*, David Bearman, ed. (Pittsburgh: Archives & Museum Informatics, 1991), p. 24. Bearman is credited as being one of the most influential postcustodial archival theorists of the late twentieth-century. See Terry Cook, "The Impact of David Bearman on Modern Archival Thinking," pp. 15-37.

[52] Terry Cook, "Electronic Records, Paper Minds: The revolution in information management and archives in the post-custodial and post-modernist era," *Archives & Manuscripts* 22 (November 1994), pp. 300-328. The term "post-custodial" was first used by American archivist F. Gerald Ham in his paper "Archival Strategies for the Post-Custodial Era," *The American Archivist* 44 (Summer 1981), pp. 207-216. See also Frank Upward & Sue McKemmish, "Somewhere Beyond Custody," *Archives & Manuscripts* 22 (May 1994), pp. 137-149. As an approach, postcustodialism is more concerned with influencing electronic recordkeeping practices in the creation environment than it is the actual archival custody of electronic materials.

[53] Bearman. *Electronic Evidence*, pp. 278-292. See also Bearman, "An Indefensible Bastion," pp. 16-24. The role of archivists within the distributed custody model is the provision of management oversight for records within the custody of their originating agency. An application of the distributed custody model is discussed in Terry Cook, "Leaving Archival Electronic Records in Institutions: Policy and Monitoring

realm of the personal were the impetus behind the first discussions of the unique challenges of archiving personal records in the electronic age.

Responding in 1994 to the distributed custody approach, Adrian Cunningham argues that due to unregulated personal recordkeeping behaviors, unstructured private creating environments, and the transitory nature of individuals, distributed custody "is not a solution that can be applied to personal records" and adds "Governments and organizations may exist for indefinite periods of time or have cooperative successor organizations. Private individuals have an unfortunate habit of dying…"[54] Framed within the model of distributed custody, the expiration of a life is also the cessation of archival preservation and access. Cunningham further explains personal records are traditionally captured when records creation has effectively ceased – that is – at the end or near end of the private creator's life, which in the electronic age of rapid hardware and software obsolescence results in the loss of usability and contextual meaning in otherwise fragmented personal collections.[55] In this passive method of records acquisition, archivists are left to deduce or speculate on the details of personal recordkeeping behaviours and infer the associated meaning and value of the records generated by those behaviours. While Cunningham criticizes the application of distributed custody to the private domain, he does see the utility of postcustodialism in the front-end acquisition of personal archives in what he refers to as "pre-custodial intervention."

Pre-custodial intervention is a liaising strategy whereby the archivist becomes actively involved in the recordkeeping processes of individuals to ensure electronic

---

Arrangements at the National Archives of Canada," *Archives and Museum Informatics* 9:2 (1995), pp. 141-149.

[54] Adrian Cunningham, "The Archival Management of Personal Records in Electronic Form: Some Suggestions," *Archives & Manuscripts* 22 (May 1994), p. 99.

[55] Ibid., pp. 94-105.

records are "properly created, managed and documented in the first instance" so archivists may capture the required content data in addition to the contextual and structural elements of the records to support long-term preservation and the provision of access.[56] This is accomplished by an adherence to standardized recordkeeping systems agreed on by both the private individual and the archivist thereby satisfying contemporary media standards, the production of adequate supporting documentation, and the early securing of an in-principle agreement diminishing the potential for unwanted protraction in the transfer of the records to the archival institution. This liaising or interventionist model also shifts the acquisition practices of archival institutions away from the end of a career or life toward a policy of targeting potential donors in the period of active records creation.[57] As a novel and rather unorthodox theoretical approach to personal archives, there are a number of objections to pre-custodial intervention that deserve consideration.

First, the temporal orientation of this liaising model presupposes the identification of those individuals within society whose lives and subsequent recordkeeping practices warrant archival intervention. Put differently, an appraisal methodology, such as the documentation strategy, precedes the implementation of any pre-custodial intervention as it aims to discern early the historical and cultural significance of individual lives in contemporary contexts. Consequently, finding the appropriate records creators which meet the criteria of a documentation strategy may prove to be too demanding on archivists and the institutions they work for. Secondly, archival intervention of this variety also suggests the provision of one-on-one guidance and assistance which, as the

[56] Adrian Cunningham, "Waiting for the Ghost Train: Strategies for Managing Personal Records Before it is Too Late," *Archival Issues* 24 (1999) p. 58.
[57] Ibid., p. 101.

work of Canadian archivist Lucie Paquet reveals, is quite labour-intensive in practice. For example, in her exploratory article on the acquisition of electronic personal records, Paquet affirms the need for archivists to become actively involved with donors in the records creation phase through on-site investigations of contemporary and legacy hardware and software systems, interviews with donors and the documentation of their recordkeeping systems, establishing relationships between electronic and analog records, and scheduling the eventual transfer of electronic records to archival institutions.[58] And finally, this interventionist approach, specifically the involvement of archivists in records creating processes, may also lead to self-conscious recordkeeping practices and "unnatural" personal archives when the individual is informed that their past, present, and future records are, in essence, already appraised as having cultural or historical value.

Criticisms aside, pre-custodial intervention represents a turning point in the literature on personal archives for two reasons: one, it further destabilizes the assumption that government recordkeeping theory is unproblematically transposed to the personal realm; and two, it successfully articulates that traditional post hoc approaches to personal archives are, as Cunningham argues, "patently inadequate in the electronic environment."[59] At the core of Cunningham's thesis is an understanding that personal electronic records will not survive to cross the archival threshold and therefore contribute to the construction of documentary heritage if archivists are unable to adapt their practices to reflect the dynamic processes of private records creation in electronic environments. In closing, pre-custodial intervention was introduced as a conceptual approach in opposition to the more strident elements of a larger post-custodial narrative

[58] Lucie Paquet, "Appraisal, Acquisition and Control of Personal Electronic Records: From Myth to Reality," *Archives & Manuscripts* 28 (November 2000), pp. 71-91.
[59] Cunningham, "Waiting for the Ghost Train," p. 59.

occurring throughout the 1990s. During this period, however, a separate records-centered approach was developed in the interest of ensuring the reliability and authenticity of electronic records through metadata discovery and preservation.

Metadata, Authenticity, and Personal Electronic Records

Metadata is often defined as "data about data", however, within the context of electronic archives it is more aptly defined as structured information about the logical characteristics of an electronic record to assist in its identification, description, and management.[60] Archival metadata such as the Dublin Core Metadata Initiative (DCMI), Encoded Archival Description (EAD), and Encoded Archival Context (EAC) are all based on syntactical rules, such as extensible markup language (XML), and semantics or the meaning of those rules within metadata schema.[61] Metadata schema are standardized "sets of metadata elements designed for a specific purpose" which may be further optimized in metadata application profiles (Figure 1.1.b) for a particular domain.[62] The extent of how much information metadata provides about a given record is based on its level of granularity, with low granularity (Figure 1.1.a) providing basic "tombstone" information, and high granularity (Figure 1.1.b) providing more robust contextual information about electronic records. These levels of granularity are determined by the

---

[60] In order to render the arrayed binary data of electronic records intelligible for human interpretation, internal structural metadata is generated to indicate how that record is built and specifies, for example, how it is to be manipulated by application software such as a word processor. Internal structural and administrative metadata is generated by programmers and developers involved in the construction of electronic materials or through automated processes with secondary creators generating additional descriptive and administrative metadata.

[61] Other archival metadata schema include: Metadata Object and Description Schema (MODS) and Metadata Encoding and Transmission Standard (METS).

[62] National Information Standards Organization (NISO), *Understanding Metadata*, (Maryland: NISO Press, 2004), Available at www.niso.org/publications/press/UnderstandingMetadata.pdf (accessed 15 May 2011). Rachel Heery and Manjula Patel, "Application profiles: mixing and matching metadata schemas," *Ariadne* 25 (September 2000). Available at http://www.ariadne.ac.uk/issue25/app-profiles/ (accessed 15 May 2011). Figure 1.1.b. is a metadata application profile based on the Dublin Core, MARC Relator Codes, Collection description Terms, General Description Terms, and Collection Type Vocabulary Terms schema.

number of mandatory elements or attribute fields populated either manually by the creator, or automatically via software application or electronic management system. Without metadata, records are simply “data” devoid of structure and contextual meaning. Furthermore, the successful acquisition and exploitation of metadata is essential to the long-term preservation of electronic records and to the establishment of electronic archives. Ensuring the longevity of electronic records and their respective metadata is the mandate of the International Research on Permanent Authentic Records in Electronic Systems (InterPARES) Project.

Figure 1.1.a DCMI Element Set[63]

<dcterms:contributor>
</dcterms:contributor>
<dcterms:coverage> </dcterms:coverage>
<dcterms:creator> </dcterms:creator>
<dcterms:date> </dcterms:date>
<dcterms:description>
</dcterms:description>
<dcterms:format> </dcterms:format>
<dcterms:identifier> </dcterms:identifier>
<dcterms:language> </dcterms:language>
<dcterms:publisher> </dcterms:publisher>
<dcterms:relation> </dcterms:relation>
<dcterms:rights> </dcterms:rights>
<dcterms:source> </dcterms:source>
<dcterms:subject> </dcterms:subject>
<dcterms:title> </dcterms:title>
<dcterms:type> </dcterms:type>

Figure 1.1.b DCMI Application Profile[64]

<dc:identifier></dc:identifier>
<dc:title></dc:title>
<dcterms:alternative></dcterms:alternative>
<dcterms:abstract></dcterms:abstract>
<dc:format></dc:format>
<dcterms:extent></dcterms:extent>
<dc:language></dc:language>
<dc:type></dc:type>
<dc:rights></dc:rights>
<dcterms:accessRights></dcterms:accessRights>
<cld:accrualMethod></cld:accrualMethod>
<cld:accrualPeriodicity></cld:accrualPeriodicity>
<cld:accrualPolicy></cld:accrualPolicy>
<dcterms:provenance></dcterms:provenance>
<dcterms:audience></dcterms:audience>
<cld:logo></cld:logo>
<dc:subject></dc:subject>
<dcterms:spatial></dcterms:spatial>
<dcterms:temporal></dcterms:temporal>
<dcterms:created></dcterms:created>
<cld:dateContentsCreated></cld:dateContentsCreated>
<dc:creator></dc:creator>
<marcrel:own></marcrel:own>
<gen:isLocatedAt></gen:isLocatedAt>
<gen:isAvailableVia></gen:isAvailableVia>
<dcterms:isReferencedBy></dcterms:isReferencedBy>
<dcterms:isPartOf></dcterms:isPartOf>
<dcterms:hasPart></dcterms:hasPart>
<dc:description></dc:description>
<dc:relation></dc:relation>

In the early 1990s, archival educator Luciana Duranti initiated the adaptation of diplomatics, a seventeenth-century analytical-evaluative technique, to determine the authenticity of electronic records.[65] Diplomatics is defined as "the analysis of the genesis, inner constitution and transmission of documents, and of their relationship with the facts represented in them and with their creators" where the document as an expression of

[63] Table based on Dublin Core Metadata Initiative, *Dublin Core Metadata Element Set, Version 1.1*. Available at http://dublincore.org/documents/dces/ (accessed 15 May 2011).
[64] Table based on Dublin Core Metadata Initiative, *Dublin Core Collection Description Application Profile*. Available at http://dublincore.org/groups/collections/collection-application-profile/2004-08-20/#dcidentifier (accessed 15 May 2011).
[65] Yvette Hackett, "The Search for Authenticity in Electronic Records," *The Moving Image 3*, no.2 (2003), pp. 100-107.

ideas (content) is transmitted or described according to both physical and intellectual rules of representation (form).[66] Modern diplomatics is the rigorous interrogation of the internal and external elements of individual electronic documents to obtain important information on their context of creation and use, but ultimately does so in the interest of determining and maintaining their authenticity.

Diplomatic analysis of electronic records is the foundation of the InterPARES Project and its corresponding Authenticity Task Force.[67] Unlike post-custodial theory, InterPARES principally examines the electronic record from an archives-as-place perspective: "For the transparency of its preservation, its security and its stability, it is necessary that the record pass the archival threshold, the space beyond which no alteration of permutation is possible and where every written act can be treated as evidence and memory."[68] InterPARES is a series of investigative projects seeking to develop conceptual requirements for the long-term preservation of authentic records created in electronic form where an authentic record is defined as one that "can be proven to be (i) what it claims to be and (ii) free of falsification or inappropriate modification."[69]

---

[66] Luciana Duranti, "Diplomatics: New Uses for an Old Science," *Archivaria* 28 (Summer 1989), pp. 7-27. In her analysis of the diplomatic method, Susan E. Storch notes "When applied to specific documents, diplomatic analysis reveals 'their genetic process, their creators, the facts they record, and the way in which their form reflects their function and the activities of their creators.'" Diplomatics: Modern Archival Method or Medieval Artifact?" *The American Archivist* 61 (Fall 1998), p. 367.

[67] The history of InterPARES begins in the 1994 Preservation of the Integrity of Electronic Records, or "UBC Project" at the University of British Columbia whose findings led to the initiation of InterPARES I (1999-2001), the subsequent InterPARES II (2002-2007) and the currently ongoing InterPARES III (2007-2012).

[68] Luciana Duranti, "Archives as a Place," *Archives & Manuscripts* 24 (1996), p. 252.

[69] Heather MacNeil, "Providing Grounds for Trust: Developing Conceptual Requirements for the Long-Term Preservation of Authentic Electronic Records," *Archivaria* 54 (Fall 2002), p. 53. Within this conceptual framework, the authenticity of a record is assessed "in relation to its identity (i.e., was it written by the person who purports to have written it?) and its integrity (i.e., has it been altered in any way since it was first created and, if so, has such alteration changed its essential character?)" See also InterPARES 1, Authenticity Task Force, *Requirements for Assessing and Maintaining the Authenticity of Electronic Records* (March 2002): pp. 1–11. Available at http://www.interpares.org/book/interpares_book_k_app02.pdf (accessed 30 May 2011).

Like the Pittsburgh Project, InterPARES focused on the institutional electronic record, reflecting the heavy emphasis on institutional records in the approach to diplomatics adopted by Duranti. The formal criteria developed by InterPARES for the assessment and maintenance of authenticity are divided into two key areas: the benchmark requirements supporting the presumption of authenticity, which is "an inference that is drawn from known facts about the manner in which a record has been created and maintained"; and the baseline requirements supporting the production of authentic copies of electronic records.[70] The former requirements relate to the activities of the records creator in the creation environment, while the latter requirements relate to the activities of the records preserver or archivist. Discovering and supporting authenticity within the context of these benchmark and baseline requirements is dependent on one key element: the creation, verification, and preservation of metadata.

The Presumption of Authenticity of Electronic Personal Archives

In 2006, the InterPARES 2 Project published a set of guidelines for individual records creators which are introduced by the following statement:

> These guidelines have been developed for individuals who create digital materials in the course of their professional and personal activities to help them make informed decisions about making and maintaining these materials in ways that will help ensure their preservation for as long as they are needed.[71]

Seeking to extend the applicability of InterPARES 1 requirements to records creating environments outside of administrative and legal domains, InterPARES 2 focused on records produced in complex digital environments in the course of artistic, scientific, and governmental activities and developed guidelines for making and maintaining digital

---

[70] InterPARES 1, Authenticity Task Force, *Requirements for Assessing and Maintaining the Authenticity of Electronic Records*, p. 3.
[71] InterPARES 2 Project, *Creator Guidelines: Making and Maintaining Digital Materials: Guidelines for Individuals,* (Vancouver: InterPARES Project, 2006).

materials. The InterPARES 2 *Creator Guidelines* encourage the creation of highly granular or detailed metadata for electronic documents created by a private individual, groups of individuals, or small organizations to "help ensure that records that merit long-term preservation in an archival repository will become accessible when they are turned over to the care of a trusted custodian."[72] What is more, it is recommended that each of the element attributes in these guidelines be explicitly expressed and inextricably linked to every record generated within the active creating environment. The concluding section of this chapter of the thesis examines the utility of the InterPARES 2 *Creator Guidelines*, which are visually represented in Figure 1.2.

[72] Ibid.

Figure 1.2 InterPARES Creator Guidelines: Identity and Integrity Metadata Requirements[73]

1. Identity of the record:

A. Names of the persons involved in the creation of the digital materials:
- the author
- the writer
- the originator
- the addressee
- the recipient

B. Name of action or matter
C. Documentary form
D. Digital presentation
E. Date(s) of creation and transmission:
- chronological date
- dates of transmission and/or receipt
- archival of filing date

F. Expression of documentary context
G. Indication of attachments
H. Indication of copyright or other intellectual rights
I. Indication of the presence or removal of a digital signature
J. Indication of other forms of authentication
K. Indication of draft or version number
L. Existence and location of duplicate materials outside the digital system

2. Integrity of the record:

A. Name of handling person/office
B. Name of person or office with primary responsibility for keeping the materials
C. Indication of annotations added to the materials
D. Indication of any technical changes to the materials or to the application(s) responsible for managing and providing access to the materials
E. Access restriction code
F. Access privilege code
G. Vital record code
H. Planned disposition

A recent case study reveals that supporting and maintaining the presumption of authenticity of private electronic records is not only a laborious task but also one fraught with complexity. Seeking to determine if the InterPARES benchmark requirements can

[73] Figure based on InterPARES 2 Project, *Creator Guidelines: Making and Maintaining Digital Materials: Guidelines for Individuals*. These Creator Guidelines are derived from the InterPARES 1: Authenticity Task Force, *Benchmark Requirements Supporting the Presumption of Authenticity of Electronic Records* which are "the conditions that serve as a basis for the preserver's assessment of the authenticity of the creator's electronic records. Satisfaction of these benchmark requirements will enable the preserver to infer a record's authenticity on the basis of the manner in which the records have been created, handled, and maintained by the creator." The archivist may presume the authenticity of electronic personal records based on the number of Identity and Integrity requirements that have been met and the degree to which each has been met. "The requirements are, therefore, cumulative: the higher the number of satisfied requirements, and the greater the degree to which an individual requirement has been satisfied, the stronger the presumption of authenticity."

serve as a standard for assessing the authenticity of electronic personal collections, archivist for personal manuscripts at Yale University Michael Forstrom asserts:

> An individual author's electronic records are rarely, or at least less often, subject to the juridical-administrative requirements governing institutional records. Further, to the extent that the records reflect the work of an individual outside an institutional or networked environment, there may be few or no transactional attributes linked to the records.[74]

Acknowledging that many electronic records within private manuscript collections forgo "pre-custodial intervention" or any appraisal prior to acquisition, Forstrom discovered that many of the requirements based on the transactional attributes of records could not be demonstrated or established in the electronic records of individuals and found the InterPARES benchmark requirements to be "irrelevant or impractical."[75] With this, Forstrom dedicates the remainder of his case study to testing the baseline requirements and concludes that the authenticity of electronic personal archives is "a challenge to assess and guarantee when we know little or nothing about the manner in which records have been created and maintained."[76]

An important issue not addressed in Forstrom's case study, or in the *Creator Guidelines* for that matter, is how and where a private individual would document this required metadata. Presumably, the private records creator would link metadata schema (Figure.1.1a or 1.1.b) based on the syntactic and semantic elements of the *Creator Guidelines* (Figure 1.2) to each record, or group of records, and populate the element fields manually or through automated processes. However, the actual manifestation of this metadata presupposes the implementation of front-end technology, such as Electronic

---

[74] Michael Forstrom, "Managing Electronic Records in Manuscript Collections: A Case Study from the Beinecke Rare Book and Manuscript Library," *The American Archivist* 72 (Fall/Winter 2009), p. 465.
[75] Ibid., pp. 461 & 466. Forstrom suggests that the benchmark requirements be modified or dropped altogether in favour of other criteria.
[76] Ibid., p. 472.

Records Management System (ERMS) software, which is deployed in corporate and government settings to regulate the management of electronic records and their associated metadata from creation to final disposition. However, the deployment of ERMS within personal digital environments is unlikely to be met with much success as this technology, much like the ideas behind its development, is predicated on organizational accountability, business functions and activities, as well as government or corporate employee compliance with the system. Furthermore, there is also a low-probability that the human-mediated assignment of metadata to personal electronic records could support even a "reasonable" presumption of authenticity unless mechanisms for creating the required metadata are built into the design of private sector commercial software applications, which to date has not occurred. Without automatically assisted metadata functionality, personal records creators are thus required to populate multiple metadata fields manually for each individual record.[77]

In its scale and scope, the InterPARES project remains to date one of the largest electronic records research initiatives and has had considerable influence on theory and practice in both the archival and records management landscapes. InterPARES has brought to the fore the importance of maintaining authentic electronic records in an age when information is so easily corrupted, lost, and deleted in both creation and preservation environments, and has provided the archival standard for the long-term preservation of trustworthy and reliable electronic records for the purposes of accountability. Moreover, the InterPARES *Creator Guidelines* make explicit the complex nature of the relationship between personal records creators and archivists in the

[77] A question that may be asked is: do private individuals place such a high premium on personal accountability as to motivate them enough to ensure the authenticity of the electronic records they create throughout their everyday activities?

electronic age and demonstrate the importance of early metadata creation as, from the example provided by Forstrom, it is exceedingly difficult to discover important contextual information about electronic records if they are not accompanied by adequate metadata.

Summary of Chapter

Theoretical perspectives on personal archives have developed in reaction to archival principles predicated on institutional records and archival strategies deployed in government settings. As the literature shows, archivists working within the personal-private dimensions of the profession have strived to create value taxonomies unique to the records of private individuals, while at the same time largely defining personal archives through an articulation of what they are not – that is – the product of business or government functions. Indeed, a great deal of scholarship reviewed in this chapter has focused on the reconciliation of personal and institutional archival principles. The institutional and ideological divisions highlighted in the literature reviewed here are ultimately due to differing interpretations of evidence, recordkeeping, appraisal value, and the role of the archivist in constructing documentary heritage from both public and private sources.

Electronic records have challenged many of the assumptions and approaches around the acquisition and preservation of personal records as traditional passive archival mediation of these sources fails to operate effectively when the duration of stability in documentary forms has been drastically reduced. Efforts to conceptually re-structure the electronic personal records creating environment in the image of accountable institutions have failed for many practical reasons, but fundamentally so because of a basic human

instinct to resist the imposition of standards, rules, and restrictions on imagination, creative impulse, self-representation, and the individual right to remember and the right to forget. Moreover, while all archivists desire to establish and maintain the true provenance and authenticity of electronic personal records, they have yet to ask themselves if these records need to be held to the same rigorous standards as those records used in corporate, legal, and government settings. Depending on the conclusion at which archivists arrive, a significant amount of work may need to be done in order to translate highly conceptual and idealized metadata requirements into actual standards that can be reasonably met by the everyday physics of electronic records creation in the personal realm.

Throughout this chapter, archivists have expressed a critical need to better understand the complexities of the pre-custodial environment, for it is in this personal domain that rich provenancial information such as personal recordkeeping activities and cultures, motivations, behaviors, and designations of values are performed, manifested, and realized. Yet, much archival theorizing on the pre-custodial activities of individuals has been based on what archivists have come to glean from collections of personal diaries, letters, photographs, and manuscripts created before the digital age and while there are similarities between personal paper and electronic records, there are also many differences. As archivist Joan Schwartz aptly notes "we must first acknowledge that the 'new' media configure not only old information in new ways, but also different information in previously unimaginable ways." [78] The next chapter of this thesis

[78] Joan M. Schwartz, "'We make our tools and our tools make us': Lessons from Photographs for the Practice, Politics, and Poetics of Diplomatics," *Archivaria* 40 (Fall 1995), p. 41.

examines personalized digital archiving environments from a largely non-archival perspective.

CHAPTER TWO

PERSONALIZED DIGITAL ARCHIVING ENVIRONMENTS

> The personal archive of a living person is, of course, a dynamic entity: a 'living archive' with new objects being created, others being acquired, amended, and discarded. [1]
>
> Inscribers and pre-archival custodians of records document some things and not others (that is an appraisal decision of sorts) and they choose to destroy certain records, without knowledge of archives, or offer only certain records to archives, holding back others for other times.[2]

The phrase "archives in the wild", as defined by British archival curator Jeremy Leighton John, refers to "the personal digital archives that exist outside an official long-term repository" including the personal archives of academics, literary and political figures but also the digital collections of ordinary, everyday people.[3] These archives in the wild are created and preserved by individuals with diverse recordkeeping behaviours and are comprised of dynamic documentary forms dispersed throughout incalculable online and offline digital landscapes. Within this wilderness, digital archives are usually created, accumulated, and maintained instinctively and expediently as opposed to systematically and routinely as they are in institutional environments. More, this digital wilderness is the hinterland for memory institutions (archives, libraries and museums) whose mandate is the acquisition, preservation, and provision of access to collections of personal records. Yet, micro-level analyses of the context of creation and use of records in private digital environments remains a much neglected area of study and is in many

---

[1] Jeremy Leighton John et al., "Digital Lives, Personal Digital Archives for the 21st Century: An Initial Synthesis, Beta Version 0.2," (March 2010). Available at http://britishlibrary.typepad.co.uk/files/digital-lives-synthesis02-1.pdf (accessed 21 April 2010). p. 9. Jeremy Leighton John is principal investigator of the Digital Lives Research Project and curator of eMANUSCRIPTS at the British Library.

[2] Tom Nesmith, "Reopening Archives: Bringing New Contextualities into Archival Theory and Practice," *Archivaria* 60 (Fall 2005), pp. 263-264

[3] Leighton John et al., "Digital Lives," p. 5.

respects still frontier research in professional archival scholarship. Moreover, there is a consensus among archivists that "only with an accurate and comprehensive perception of how electronic records are organized can archivists have a good understanding of what they are going to deal with at the time that records are transferred to archival systems."[4] Chapter two of this thesis examines personal digital recordkeeping strategies, appraisal and value, and digital preservation practices from the perspective of Personal Information Management (PIM) studies to discover why, how, and where individuals create and preserve documentary forms in the digital era.

Personal Information Management (PIM)

> We all keep information in our work and domestic lives. It may be books, notes, diaries, personal records, files or whatever. This is personal information not necessarily in the sense that it is private, but that we have it for our own use. We own it, and would feel deprived if it were taken away.[5]

Personal information management or PIM is both the practice and the study of the activities people perform to acquire, organize, maintain, and retrieve digital information for everyday use in personal computing environments. PIM seeks to discover innovative ways to assist individuals in managing excessive volumes of digital information more efficiently through the design of successful software and hardware to meet specific objectives in what may referred to as PIM technologies or applications.[6] The development of PIM applications is in most cases commercially driven to assist individuals in the short-term management of a particular technological format such as

---

[4] Jane Zhang, "The Principle of Original Order & The Organization and Representation of Digital Archives," Faculty of the Simmons College Graduate School of Library and Information Science, PhD Dissertation, (2010), p. 189.

[5] M.W. Lansdale, "The psychology of personal information management," *Applied Ergonomics* 19:1 (March 1988), p. 55. This article is commonly cited as the first expression of the term "personal information management" as a practice and area of study.

[6] William Jones and Jaime Teevan, eds., *Personal Information Management* (Seattle: University of Washington Press, 2007), pp. 6-17.

email, text, image and music files, or Web bookmarks. For example, ubiquitous PIM applications include email client software (Microsoft Office Outlook and Mozilla Thunderbird), file manger applications (Windows Explorer and Mac Finder), as well as organizational calendars (Google Calendar) and music file managers (iTunes). Emerging and more robust PIM applications include note-taking software suites (Microsoft OneNote and Evernote), web-based file managing services (DropBox and iCloud), as well as reference/personal bibliographic management software (Zotero and Mendeley). PIM applications may be proprietary or open-source, used online or offline, and synchronized between desktop (Mac OSX) and mobile (iOS) operating systems.

While they continue to evolve over time, all PIM technologies invariably involve four primary functions or mechanisms to create, arrange, remove, and access information in personal digital collections.[7] However, archivally oriented PIM adds a fifth function which may be referred to as the long-term preservation of personal digital information. Archival PIM considers the factors involved in maintaining personal digital information throughout its entire lifecycle and "is directed at securing authentic personal digital objects and making them readily available for use and reuse by the individual creators and owners beyond the immediate future."[8] In its endeavors to design better information management technology, PIM research is obligated to examine and evaluate user behaviours and strategies involved in individual recordkeeping, appraisal, and preservation activities and as such provides unique insights and an alternative perspective on personal digital archives. To truly articulate the significance of PIM research to the

[7] Richard Boardman, "Improving Tool Support for Personal Information Management," Department of Electrical and Electronic Engineering, Imperial College London, PhD Dissertation, (2004), pp. 11-24.
[8] Leighton John et al., "Digital Lives," p. x.

archival profession, however, requires that it first be situated within current archival theory and methodology.

The Pre-Custodial Environment

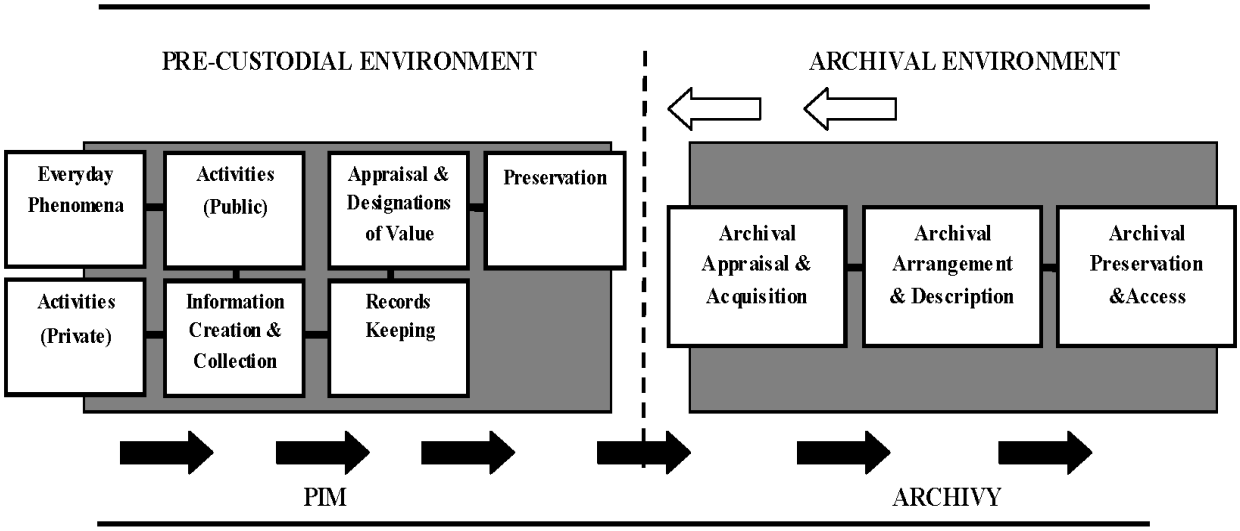
Archivists principally adhere to one of two conceptual models, or derivatives thereof, when discussing the creation and management of documentary forms: the records life cycle and the records continuum. The life-cycle concept portrays records traversing two phases with eight particular stages: a records management phase consisting of stages relating to creation, classification, maintenance and use, and disposition, succeeded by an archival phase with stages of selection and acquisition, description, preservation, and use.[9] Records continuum thinking on the other hand, posits the passing of records through four integrated time-space dimensions involving creation, capture, organization, and pluralization. While there are both similarities and differences between the records life cycle and the records continuum, both models concede a temporal and spatial period of documentary activity preceding the archival mediation of records which may be referred to as the pre-custodial environment.[10] Archivists ascertain what they can about this pre-custodial environment, such as the circumstances surrounding the creation and use of personal records and the relationships between them in their endeavours to determine provenance in retrospect. In personal archives these extrapolations are in essence based

[9] Jay Atherton, "From Life Cycle to Continuum: Some Thoughts on the Records Management – Archives Relationship," *Archivaria* 21 (Winter 1985-86), pp. 43-51. Atherton amended the traditional life cycle model from eight stages to a single-phase, four-stage model consisting of creation, classification, scheduling, and maintenance and use. In doing so, Atherton bridges the gap between records management and archival management.

[10] Barbara Reed, "Reading the Records Continuum: Interpretations and Explorations," Available at www.records.com.au/pdf/Reading_the_Records_Continuum.pdf (accessed 11 June 2011). Originally published in *Archives & Manuscripts* 33:1 (2005). In the Australian records continuum, this documentary activity occurs within the first (create) and second (capture) dimensions, or as Reed explains, "the first 'create' dimension of the records continuum represents the locus of all action…including representations of actions in documents…where characteristics from the second dimension, records, now attest to evidence of action."

on the documentary fragments the archivist possesses at the point of acquisition which, depending on the creator or donor of the records, may be quite limited in quantity and quality. PIM studies, on the other hand, document extant personal recordkeeping systems *in situ*.

Figure 2.1 Temporal and Spatial Disparity in Personal Records Mediation



The traditional records life cycle model definitively hinges on institutional transactions and is therefore inclined to express the trajectory of documentary forms through fixed-linear stages of records management. More, this model is often transposed, either by implication or by default, to personal archives in both digital and non-digital domains. As a result, pre-custodial creation, recordkeeping, appraisal, and preservation of personal documentary forms are compartmentalized as a single epoch (creation) in a finite series of temporal and spatial business-like progressions (maintenance, scheduling, and disposition) leading to the record's eventual mediation by archivists. These progressions through time and space inform the basic provenance of records which is in turn either confirmed or denied by archival mediation. In personal-private settings, it may be more

appropriate to re-conceptualize the records life cycle as occurrences of "social and technical processes of inscription, transmission, contextualization, and interpretation" streaming through both pre-custodial and archival sites of mediation and culminating in the construction of provenancial attributes.[11] The attestation of this rich provenancial information, however, has become overtly speculative with the profoundly ephemeral and fragile nature of personal digital record. Thus, the (re)encounter of archival provenance in the digital era makes necessary the timely documentation of records creation, use, and transmission within the pre-custodial environment; a task to which the PIM discipline is aptly suited due to its temporo-spatial orientation within the mediation stream of personal records.

Personal Records Creation and Recordkeeping

> What characterises the recordkeeping behaviour of individuals and what factors condition that behaviour? What range of 'personal recordkeeping cultures' can be identified?[12]

Throughout the course of a lifetime a person will naturally document their private and public activities throughout encounters with everyday phenomena. At times this documentation is an imperative for financial reasons, required to demonstrate accountability, or essential in the performance of ongoing occupational duties. Alternately, people will compose electronic missives and narrative, take digital photos, and record audio and video without preparation or for tentative purposes. Day after day people continually and consciously document their existence from the innate need to communicate with others and to engage in life's pursuits. These inaugural acts of documentation may be viewed as the first horizons of personal recordkeeping where

[11] Nesmith, "Reopening Archives," pp. 262-263.
[12] Sue McKemmish, "Evidence of me…," *Archives & Manuscripts* 24:1 (May 1996), p. 29.

digital information is created but has yet to undergo processes of organization and management.

When personal documentation takes digital form it is at once designated as a specific file type. For instance, when an individual is creating a Microsoft Word document, data is allocated to random access memory (RAM) until an auto-save or manual save command is executed at which time the data is encoded to a unique formatting algorithm and assigned a suffix with a three to four character extension (.doc or .docx) and stored on the hard disk drive (HDD). Operating Systems (Windows-based) use these file extensions to locate and execute the associated application software (Microsoft Word) required to render the data in a specific format to make it human readable.[13] While it is extremely difficult to calculate an exact number of file formats existing in personal computing environments, it is possible however to broadly categorize them in two classes: files created *offline* such as those analogous to paper-analog forms (text, audio, video, images, and worksheets); and *online* files used exclusively in the construction of Web content (HTML, JavaScript, Cascading Style Sheets). Both classes of files rely on associated software applications while online files introduce the additional requirement of host servers to publish and manipulate content on the Web. In short, the file is the rudimentary element of personal digital records.

Many of these file classes are created and/or manipulated, at one point or another, in a common domain -- the personal computer. For example, files created and stored in a digital camera are typically edited by software such as Adobe Photoshop installed on the

[13] Windows operating system XP used the .doc extension for MS Word documents. This has since changed with the Vista OS where the suffix .docx is used. As there are thousands of proprietary and open software applications available to the public, technical registries such as FileInfo.com and PRONOM aim to provide definitive information on the thousands of file formats and extension suffixes associated with those software applications.

personal computer where they subsequently remain until further action is taken. Similarly, HTML, Cascading Style Sheets, and JavaScript files are created using text-editors installed on the personal computer before FTP and hosting servers render and make the content available on the Web. Likewise, content uploaded to file sharing platforms such as YouTube and Flickr, or social networking services like Facebook emanates primarily from files created or edited on the personal computer.[14] Regardless of its operating system (Mac OSX, Linux, Microsoft Windows, or Unix), the personal computer acts as a central hub for both the creation of digital files as well as their continued use and management in what may be referred to as series of recordkeeping actions:

> [F]iling and finding are such basic aspects of working with computers that we scarcely notice their existence – hence the lack of research – every computer user spends time and effort filing and finding every time the computer is used. As designers we should be concerned with optimizing finding and filing.[15]

All personal computers provide the ability to create folders via file manager applications (Windows Explorer or Mac Finder) to organize text, audio, video, spreadsheets, and dynamic presentations among other file classes. These folders are in turn nested hierarchically to create either deep or shallow structures containing hundreds if not thousands of aggregated digital items. As personal computers are often the centre of creation and organization, file management is a core component of personal recordkeeping. A second important hub of personal recordkeeping is email. Considered

---

[14] As with all content published on the Web, social networking and file sharing services involve server-side augmentation which client-side entities have little or no control over in terms of page layout or functionality. Smartphone technology (iPhone and BlackBerry) also supports mobile uploading of content to file sharing and social networking sites but mobile data, such as captured images, is still kept in program or application directories of the personal computer when it is synchronized with the smartphone or exported using extraction software.

[15] Deborah Barreau and Bonnie A. Nardi, "Finding and Reminding: File Organization from the Desktop," *ACM SIGCHI Bulletin* 27:3 (July 1995), p. 39.

the digital counterpart of handwritten and typewritten correspondence, email is one of the most widely used applications of computer mediated communication due to the negligible learning effort to be able to use it and its almost universal availability.[16] Like written correspondence, email is a recorded manifestation of asynchronous communication between two or more people and is in many respects the major means of non-face-to-face communication in the early twenty-first-century.[17] But far from being simple accumulations of messages, email also serves as a primary information conduit through which individuals manage daily tasks and exchange a number of different file classes through attachments or embedding.[18] As a PIM application, email has grown to encompass the functions of personal recordkeeping to such a degree that many people "tend to live in email" as demonstrated by the sheer amount of time spent using it.[19] The next section of this chapter discusses personal recordkeeping within the context of file and email management.

In her 2009 study, Sarah Henderson profiles the recordkeeping behaviours of one-hundred-twenty-five knowledge workers through interviews, surveys, and on-site

---

[16] Simon Scerri, Siegfried Handschuh, and Stefan Decker, "Semantic Email as a Communication Medium for the Social Semantic Desktop," *ESWC'08 Proceedings of the 5th European semantic web conference on the semantic web: research and applications*, Available at http://portal.acm.org/citation.cfm?id=1789410 (accessed 20 May 2011). pp. 124-125. Unlike paper-based correspondence, email adheres to a number of protocols (SMTP, POP and IMAP) in the request-response function between an email client and server. It may be argued, however, that written correspondence adheres to protocols or etiquette in, for example, letters of condolence or invitation. But this type of protocol is more culturally imposed and socially expected than it is required for the actual transmission of information.

[17] Nicholas Ducheneaut and Victoria Bellotti, "Email as a habitat: An exploration of embedded personal information management," *ACM Interactions* 8:5 (September/October 2001), pp. 30-38.

[18] Ibid., p. 70. The ubiquity of email began with the debut of personal email accounts in the mid-1990s, such as Hotmail and Yahoo!, which offered free but limited disk storage space on central servers. With the advent of Google Mail (Gmail) in 2004 and its one gigabyte (one-thousand megabytes) capacity of free storage per user, email accounts have quickly came to act more as personal repositories of information than simple inboxes where individuals are rarely forced to delete messages as a result of limited storage space. By 2009, the per-user free storage space had grown to 7.3 gigabytes.

[19] Steve Whittaker, Victoria Bellotti, and Jacek Gwizdka, "Email in personal information management," *Communications of the ACM - Personal information management* 49:1 (January 2006), p. 68.

demonstrations of a participant's document management practices.[20] From her data, Henderson determines three distinct strategies. In the first strategy, individuals construct moderate folder structures through periodic cleanups or when accumulations of documents warrant them. As the folder structure is of medium depth, browsing as opposed to executing a text-search query is employed to locate documents (*Filing*). A second identified strategy involves relatively disorganized clusters of documents where low numbers of folders are created and unsystematically used. In this second strategy, the computer desktop is the primary area of organization and can therefore be browsed with relative ease while information haphazardly filed away is retrieved by text-search query (*Piling*). The final strategy demonstrates organized folders with low numbers of unclassified documents, where individuals often preempt the influx of documents through the creation of folder categories in advance. In this strategy, individuals browse to re-find documents but rely on context, such as parent folders, to locate information within their deep folder structures (*Structuring*).

In their study of "finding" and "reminding," Deborah Barreau and Bonnie Nardi report on the collated findings of two separate investigations of the methods employed by managers, graphic artists, programmers, administrative assistants, and librarians to organize and find files on their personal computers. Barreau and Nardi highlight two basic recordkeeping strategies: location-based finding and logical (text-search query) finding.[21] They conclude that the deployment of these two strategies correlates with the type of electronic information the user is working with, which they classify as

---

[20] Sarah Henderson, "Personal Document Management Strategies," *CHINZ '09 Proceedings of the 10th International Conference NZ Chapter of the ACM's Special Interest Group on Human-Computer Interaction*, (2009). Available at http://sarahhenderson.info/tags/personal-document-management/ (accessed 11 June 2011), pp. 69-76.

[21] Barreau and Nardi, "Finding and Reminding", pp. 39-43.

"ephemeral," "working," or "archived."[22] Study participants demonstrated that ephemeral and working items are often retrieved by location-based finding as this type of information is frequently used or serves a reminding function and therefore receives prime real-estate on the computer desktop. Archived items of less immediate relevance or utility, however, are rarely organized in a systematic way and rely on text-search queries for retrieval. Barreau and Nardi conclude that while individuals keep archived information for extended periods of time, study participants indicated that selecting and establishing logical filing schemes of keywords and carefully built structures for this information often failed.[23] Users, the authors suggest, "prefer filing by location because it aids in helping them find what they need as well as serving a crucial reminding function."[24] Accordingly, location-based storage, on the desktop for instance, "assumes a small information collection (basically what the user can remember) and does not scale to large archived collections" comprised of, for example, numerous folders and multiple files.[25] In short, the way individuals use their information "is a primary determinant of how it will be organized, stored, and retrieved in the personal workspace" and as archived collections of information are "often needed in a context that is different from the one in which it was created" or in other words "secondary" to ephemeral and working information, they are spatially separated as a unique type of information.[26]

---

[22] Ibid., pp. 40-43. These three types of information are further defined as: ephemeral information (short shelf life), working information (shelf-life of weeks or months), and archived information (shelf life of months or years).
[23] Ibid., p. 42.
[24] Ibid.
[25] Scott Fertig, Eric Freeman and David Gelernter, "Finding and Reminding Reconsidered," *ACM SIGCHI Bulletin* 28:1 (January 1996), p. 67.
[26] Barbara H. Kwasnik, "How a personal document's intended use or purpose affects its classification in an office," in *Proceedings of the ACM-SIGIR 12th Annual International Conference on Research and Development in Information Retrieval* (June 1989), p. 210. Available at

Profiling the email use of sixty professional office workers, Wendy Mackay finds that individuals typically view email as a tool for task management or information management. Those viewing email as a mechanism to support a variety of time and task management activities often do not read all incoming messages, limit the number of times they read email per day, and reduce message volume by unsubscribing to mail lists.[27] These "prioritizers" run contrary to those users who attempt to read all incoming mail, save a large percentage of messages, and maintain many email folders where messages are transferred on an ad hoc basis. These "archivers" she notes, seek to keep rather than delete, view email as a means for supporting information management, and are subsequently overwhelmed by messages in their inboxes. Mackay's study reveals how the email organizational patterns of individuals are influenced by how they think about the functions of email; those who view email as an information store typically retain all of their messages and do not view the deletion of messages "as particularly useful."[28]

In their seminal exploration of email management, Steve Whittaker and Candace Sidner state:

> Email is one of the most successful computer applications yet devised. Our empirical data show however, that although email was originally designed as a *communications* application, it is now being used for additional functions, that it was not designed for, such as *task management* and *personal archiving*. We call this *email overload*.[29]

---

http://portal.acm.org.proxy1.lib.umanitoba.ca/citation.cfm?doid=75334.75356 (accessed 11 June 2011), pp. 207-210. See also Fertig, Freeman, and Gelernter, pp. 67-68.

[27] Wendy E. Mackay, "More than Just a Communication System: Diversity in the Use of Electronic Mail," in *Proceedings of the 1998 ACM conference on Computer-supported cooperative work* (New York: ACM Press, 1998), pp. 344-353.

[28] Ibid., p. 350.

[29] Steve Whitaker and Candace Sidner, "Email overload, exploring personal information management of email," *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, (1996), pp. 276-283. Available at http://portal.acm.org/citation.cfm?id=238530&dl=ACM&coll=DL&CFID=27124289&CFTOKEN=32910868 (accessed 20 May 2011).

Examining the inboxes of twenty office workers, Whittaker and Sidner identify four types of email messages: those requiring the user to execute some action (to dos); messages with considerable informational content requiring thorough examination (to reads); informational messages of undetermined significance (intermediate status); and threads of asynchronous communication (ongoing correspondence).[30] The authors note that in rationalizing an overwhelming amount of these types of information, individuals tend to perform one of three recordkeeping strategies. Those who make use of email folders are categorized as users who frequently delete and archive email items ("frequent filers") whereas those users who periodically delete and archive every one to three months are categorized as ("spring cleaners").[31] The third strategy, where users neither delete nor archive email items, are categorized as "no filers." Following the deployment of these strategies, users engage in the process of maintaining a filing system, which Whittaker and Sidner state, "is a cognitively difficult task" involving the overheads of consistently creating folders, adhering to unfailing naming conventions, and remembering the definitions of folders and their contents for every transfer of an item out of the inbox (*multiple folder definitions*).[32] Also, creating new folders may not be useful if they are synonymous with pre-existing ones (*duplication*) and created folders with only one or two messages do not significantly reduce the complexity of email management,

[30] Ibid., p. 278.
[31] Ibid., pp. 280-283. Archiving is defined by Whittaker and Sinder as the process of transferring an email item from the inbox to a separate folder.
[32] Ibid., p. 279. Ten years later, two PIM researchers conducted a similar study comparing it to the research of Whittaker and Sidner and discovered a tenfold increase in email archive size as well as a significant increase in the amount of email folders. See Danyel Fisher et. al, "Revisiting Whittaker & Sidner's 'Email Overload' Ten Years Later," in *Proceedings of the 15th International Conference on Intelligent User Interfaces*, (November 2006), pp. 309-312. Available at research.microsoft.com/pubs/69394/p309-fisher.pdf (accessed 20 May 2011).

while folders with too many messages become just as unmanageable as the inbox itself (*failed folders*).[33]

Collecting longitudinal and cross-application data on personal recordkeeping strategies, Richard Boardman and Martina Sasse provide insight into how personal recordkeeping behaviour evolves over time.[34] Profiling thirty-one participants, Boardman and Sasse argue that in the creation of documents, individuals will generally file either occasionally, extensively, or upon creation or accumulation of digital information. The authors note that those participants exhibiting high organizational effort (*pro-organizing*) tend to do so across their collections of email, files, and web bookmarks, with similar overlap found across all three collections with those individuals exhibiting low-organizational tendencies (*organizing-neutral*).[35] Revisiting these classifications in the second phase of their study, Boardman and Sasse note that changes in recordkeeping tendencies are relatively subtle and tend to be adjustments to existing pro-organizing strategies rather than major changes such as an individual changing their email strategy from *no-filer* to *spring-cleaner*.[36] The authors conclude that an individual's organizational strategy is also influenced by their method of retrieval as, for example, a person relying primarily on folder-based browsing in retrieval will invest time in filing for the "cost of filing is offset by predicted benefits at retrieval time."[37]

---

[33] Whittaker and Sidner, "Email Overload," p. 280.
[34] Richard Boardman and Martina Angela Sasse, "Stuff Goes into the Computer and Doesn't Come Out: A Cross-tool Study of Personal Information Management," in *Proceedings of the SIGCHI conference on Human factors in computing systems, CHI*, 6:1(New York: ACM Press, 2004), pp. 583 – 590.
[35] Ibid., pp. 584-588. Along with email and files, Web bookmarking is the third major area of PIM research. Web browsers support ways of returning to previously viewed web content by adding urls to "Favorites" (Internet Explorer) and "Bookmarks" (Firefox). Conversely, a person may also email a url link to himself or herself or paste it in a document.
[36] Ibid., pp. 588-590. In *Phase 2* of the study, the authors tracked the evolution of recordkeeping strategies across all three digital information collections over an average period of two-hundred-eighty-six days.
[37] Ibid., p. 589.

Additional PIM studies have identified similar recordkeeping strategies including: "cleaners and keepers", "folder-less cleaners and folder-less spring-cleaners", "neat and messy", as well as "sporadic and end-of-session filers" among others.[38] While these studies somewhat differ in how they classify user behaviours, there are three prominent attributes of personal recordkeeping to be considered. First, personal digital items are either active or dormant with the former logically situated for regular re-encounters and the latter relegated to more obscure locations. As active digital items will inevitably become inactive, agglomerations of dormant files fall victim to a "poverty of attention" where they compete against a growing amount of active information for management due diligence.[39] Second, the process of re-finding personal digital items invariably relies on the execution of queries based on keyword attributes recovered from human memory (search-based system) or contextual and spatial cues encountered through browsing (location-based system). Both approaches, however, are encumbered by an individual's capacity to recall precise detailed information from the confines of human memory, or an individual's ability to maintain proficient and persistent organizational schemes within hierarchical folder structures. Third, in all recordkeeping strategies reviewed in this section, there is little mention of individuals consciously deleting digital items, which implies that dormant items are, by default, segregated from new items and kept for

---

[38] See Jacek Gwizdka, "Email Task Management Styles: The Cleaners and the Keepers," in *SIGCHI Conference on Human Factors in Computing Systems*, (Vienna: ACM Press, 2004), pp. 1235-1238. Olle Balter, "Strategies for organizing email," in *People and Computers XII: Proceedings of HCI'97*, H. Thimbley, B. O'Conaill and P.J. Thomas, eds. (London: Springer-Verlag, 1997), pp. 21-38. Thomas W. Malone, "How Do People Organize Their Desks? Implications for the Design of Office Information Systems," *ACM Transactions on Office Information Systems*, 1:1 (January 1983), pp. 99-112. David Abrams, Ron Baecker, and Mark Chignell, "Information archiving with bookmarks: personal Web space construction and organization," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (California: ACM Press, 1998), pp. 41-48.

[39] Herbert Simon, *Computers, Communications and the Public Interest*, Martin Greenberger, ed. (Baltimore: Johns Hopkins Press, 1971, pp. 40-41. Quoted in Thomas H. Davenport and John C. Beck, "The Attention Economy," *Ubiquity* (May 2001). Available at http://ubiquity.acm.org/article.cfm?id=376626 (accessed 10 October 2011).

further cognitive processing, rather than destroyed by their creators. In digital environments retention is the norm and destruction the exception for the simple reason that there is often no motive to destroy files when limits on virtual space and financial cost cease to be determining factors.

Despite the number of inert items ostensibly allocated in digital escrow, a percentage of these items is often released to ensuing phases of recordkeeping upon the fulfillment of distinctively personal disposition criteria. The next section of this chapter examines why people choose to preserve re-encountered digital items and how individuals extend the production of digital documentation to the creation of digital records.

Personal Appraisal and Notions of Value

Professional archivists engage in two types of appraisal when acquiring the records of private individuals: macro-level appraisal to determine the historical and cultural significance of a personal collection in relation to an acquisition or collection development policy; and micro-level appraisal to determine the value of documentary forms within that personal collection in order to "separate the wheat from the chaff." Librarian and Archivist of Canada, Daniel J. Caron, and Director General of Library and Archives Canada, Richard Brown, state that these appraisal activities:

> have mostly taken place within an anachronistic time and space, in some instances, four or more evolutionary stages away from the first contexts of their human agency or intention (i.e., the identification, appraisal, selection, and declaration of analog information resources is far removed from what could be called the social sequence of their original significations of situation, experience, and meaning).[40]

[40] Daniel J. Caron and Richard Brown, "The Documentary Moment in the Digital Age: Establishing New Value Propositions for Public Memory," *Archivaria* 71 (Spring 2011), p. 7.

As active participants in the construction of documentary heritage, archivists strive to inform their appraisal decisions as much as possible on "value ascribed by those contemporary to the material."[41] Yet, in digital environments the opportunities to locate and exploit conceptions of value within their temporal context dissipate at a far quicker rate than in paper-analog domains where the relative durability of physical documentary forms permit the "passage of time" to factor into the re-discovery of value.[42] For instance, while archivists have previously been able to appraise physical media (journals and diaries, literary drafts, poems, and correspondence, or photographs and audio-visual recordings) fifty years after their creation, this interval of time has significantly contracted with information technology becoming "essentially obsolete every 18 months" while digital storage media can be expected to become obsolete within no more than five years.[43] Concurrently, the passage of time between the creation of personal records and their appraisal by professional archivists is a principal site of evolving mediations of value facilitated by implicit and explicit private appraisal decisions. As such, these values and private appraisals remain indispensable contextual elements of personal digital records critical to the understanding and exposition of their provenance.

In her discussion of the long-term fate of personal digital archives, Microsoft researcher Catherine C. Marshall states:

> Why do we need to discuss the value of digital materials separately instead of bringing in best practices from physical information management? The

---

[41] Hans Booms, "Society and the Formation of a Documentary Heritage: Issues in the Appraisal of Archival Sources," *Archivaria* 24 (Summer 1987), p. 104.
[42] Caron and Brown, pp. 6-7.
[43] Terry Kuny, "A Digital Dark Ages? Challenges in the Preservation of Electronic Information," *Sixty-third IFLA Council and General Conference*, (August 1997), p. 3. Available at http://archive.ifla.org/IV/ifla63/63kuny1.pdf (accessed 15 May 2011) and Jeff Rothenberg, "Ensuring the Longevity of Digital Documents," *Scientific American* 272:1 (January 1995), pp. 42-47.

> reason is straightforward: digital belongings accumulate at a far more precipitous and unmanageable rate than physical belongings do.[44]

Marshall suggests that as individuals rapidly create and accumulate digital materials there comes a time when conscious decisions of what to save and what to delete become unavoidable in the effort to establish control over their private archives. These decisions, Marshall affirms, are candidly linked with private assessments and designations of value which are identified through examining the source of the digital item, the actions taken by or upon it, and by its disposition.[45]

For instance, a calculation of how often a digital item is used or replicated, how much time and effort went into its creation, with whom the digital item is shared, and the ability to reconstruct its source over time produces five general notions of value: demonstrated worth, creative effort, labour, emotional impact, and stability.[46] Marshall argues that in the personal digital archive, value is not wholly ascribed to items at inception, but rather accretes with use, in custody, and transmission over time. Here, Marshall leverages the tacit appraisal decisions of individuals and expresses them heuristically in an effort to create an archiving system which expedites the cognitive process of distinguishing "between items that are valuable and items that have simply accumulated" in the personal digital archive.[47]

In 2008, the Digital Lives research team reported on a series of in-depth interviews with twenty-five individuals from the fields of politics and the academic arts

---

[44] Catherine C. Marshall, Sara Bly, and Francoise Brun-Cottan, "The Long Term Fate of Our Digital Belongings: Towards a Service Model for Personal Archives," *Proceedings of Archiving 2006*, Available at http://research.microsoft.com/apps/pubs/default.aspx?id=75527 (accessed 15 May 2011).
[45] Catherine C. Marshall, "Rethinking Personal Digital Archiving Part 2: Implications for Services, Applications, and Institutions," *D-Lib Magazine* (March/April 2008). Available at http://www.dlib.org/dlib/march08/marshall/03marshall-pt2.html (accessed 15 May 2011).
[46] Marshall, Bly, and Brun-Cottan, "The Long Term Fate of Our Digital Belongings."
[47] Marshall, "Rethinking Personal Digital Archiving Part 2."

and sciences conducted in interest of determining how and why private individuals archived in the twenty-first-century. In their initial findings, the Digital Lives team concluded that both utilitarian and emotional factors are involved in private appraisal of digital items and their subsequent retention. For instance, the research team found that "quite often interviewees could not specify exactly how a document would be used later but still felt that, as long as there was a possibility that it might be of use, it was worth keeping."[48] This evaluation of digital items for potential future usage, the authors note, is coupled with the appraisal of items for their emotional-sentimental value which individuals calculate by the time and effort expended in creating the item and by contextual factors, such as personal memories, surrounding its continued use and custody. In a later study with over two-thousand completed interview responses, the Digital Lives team found the most prominent explanations offered for archiving a digital item were: as a witness to creativity, sentimental reasons and personal memory, for future reference, to share with colleagues, and in the interest of posterity.[49] Of these explanations, individuals tended to value the witnessing of creativity above all else which "might speak for creativity as a core human value and need or as self-validation." [50]

---

[48] Pete Williams et al., "Digital Lives: Report of Interviews with the Creators of Personal Digital Collections," *Ariadne* 55 (April 2008). Available at http://www.ariadne.ac.uk/issue55/williams-et-al/ (accessed 10 May 2011). "Digital Lives' is a research project focusing on personal digital collections and their relationship with research repositories. It brings together expert curators and practitioners in digital preservation, digital manuscripts, literary collections, Web archiving, history of science, and oral history from the British Library with researchers in the School of Library, Archive and Information Studies at University College London, and the Centre for Information Technology and Law at the University of Bristol."

[49] Leighton John et al., "Digital Lives, Personal Digital Archives for the 21st Century," pp. 44-45.

[50] Ibid., pp. 44-45. This information is derived from two surveys conducted by Digital Lives. One survey was directed at academics and another at members of the digital public. Sixty-three percent of academics and forty-five percent of the digital public cited "as a witness to creativity" as the main reason for archiving digital items followed by "sentimental reasons," and "personal memory" with fifteen and twenty-six percent respectively.

In a series of surveys and interviews with fifty office workers, Steve Whittaker and Julia Hirschberg of AT&T Labs-Research examine how individuals evaluate the significance of paper-based information and the motivations behind the archiving of that information.[51] Whittaker and Hirschberg concluded that individuals measure their decisions to archive against five broad criteria: reference value, legal and administrative value, immediate availability, reminder of encountered information, and distrust of external information storage.[52] However, the authors posit individuals archive information for reasons beyond routine business functionality and factor emotional and sentimental reasons into the decisions they make regarding the retention and disposition of recorded information. For example, the authors note documents such as reviews of published papers, successful research prototypes, and reference documents containing personal annotations have "little relevance for likely activities, but they [people] still cannot part with it, because it is part of their intellectual history."[53] One participant acknowledged that although they had no identifiable need for their archived papers and could not articulate why they kept them, "*Sentiment…or something*" prevented them from throwing the papers out.[54] Emotional or sentimental reasons for archiving, while not easily rationalized in business settings, are nevertheless contemplated by individuals in the appraisal of their recorded information.

The broad abstraction of "sentimental value" is deconstructed in a PIM study seeking to design supportive technologies for the preservation of analog and digital

---

[51] Steve Whittaker and Julia Hirschberg, "The character, value and management of personal archives," *ACM Transactions on Computer Human Interaction* 8 (2001), pp. 150-170. The study coincided with an office move which the authors believed would motivate their participants to make clear decisions about retention and disposition of materials.
[52] Ibid., pp. 155-156.
[53] Ibid., pp. 156-157.
[54] Ibid., pp. 156-157.

objects within familial environments. In their excavation of the home archive, Microsoft researchers David Kirk and Abigail Sellen explain sentimental value is "tied to the notion of constructing or bolstering a sense of identity, through knowing who one is by keeping hold of memories and reflections of the past" (*Constructing the persona*). [55] Study participants also informed the authors of the value of sharing a collective past, where certain objects such as photographs supported family "connectedness" or when objects are designated to be passed down through successive generations (*Connecting with a shared past*). Similarly, participants also attribute importance to objects intended for broader historical considerations by persons unknown in what may be referred to as personal or familial legacy value (*To preserve a legacy*).[56] Kirk and Sellen conclude the impulse to honour the past permeates many of reasons why people archive as, for example, photographs appraised for their utility in the construction of persona or personal memory also express reverence for others in the photograph and may serve "to elevate others into family consciousness (*In Honorium).*"[57]

A comparable study observes that individuals archive for multiple reasons and in multiple ways, yet communicate five common motivations: to store and retrieve information for later use (*Finding it later*); as a testament to personal and professional achievement (*Building a legacy*); to facilitate access by others (*Sharing resources*); out of anxiety of losing important information (*Fear of loss*); and as a reflection or expression of themselves (*Identity construction*).[58] The study also identifies a strong connection

---

[55] David Kirk and Abigail Sellen, "On Human Remains: Excavating the Home Archive," *Microsoft Technical Report* (June 2008). Available at http://research.microsoft.com/apps/pubs/default.aspx?id=70595 (accessed 15 May 2011), p. 5.
[56] Ibid., pp. 5-9.
[57] Ibid., pp. 6-7.
[58] Joseph Kaye, et al, "To Have and to Hold: Exploring the Personal Archive," *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 1-6. Available at

between these motivations and how an individual structured his or her archives. For instance, a study participant expressing anxiety over loss of information carefully archived his entire filing system in such a way as to support easy incremental backups to avoid catastrophic loss of information in a hard-drive crash.[59] Comparatively, study participants who rigorously documented life or career milestones or who kept information expressing personal achievement, did so not in the interest of future information retrieval or out of fear of loss, but rather in the interest of contributing to a grander narrative of accomplishment or to reinforce a sense of self based on perceived values of dedication and personal worth.[60] The investigators of this study argue there is an important relationship between the design of personal archives, the motivations involved in keeping them, and the value(s) one ascribes to digital materials such as value pertaining to personal legacy and value involved in the continuing construction of self over time.

---

http://jofish.com/writing/tohaveandtohold.pdf (accessed 20 May 2011). In 2006, the research team published the results of this study which involved interviews with forty-eight individuals as well as on-site visits of their personal archives.

[59] Ibid., pp. 5-9

[60] Ibid., p. 6. The archives of one study participant contained an antiquated laptop computer which had been used to create her PhD. dissertation some years previous. The authors note that the laptop failed to support ongoing work and the information within it was inaccessible, yet this digital remnant still served as a reflection of accomplishment and reinforced a sense of self based on perceived values of education and dedication.

Figure 2.2 Value Folksonomy for Personal Archives[61]

PIM Identifications of Value

Identity Value: Items involved in the continued construction and expression of the self.
Personal Memory Value: Items facilitating re-encounters with the personal past.
Personal and Familial Historical Value: Items recording individual and shared narrative.
Emotional and Sentimental Value: Items witnessing creativity and achievement, and eliciting emotion.
Functional Value: Items serving as reference or administrative utility in day-to-day life.
Posterity and Legacy Value: Items perpetuating a conception of a life or career for external consideration.

Archival Identifications of Value

Evidential Value: Records documenting functions, activities, and transactions of persons.
Informational Value: Records providing information on significant persons and of potential use for historical and sociological research.
Narrative Value: Records involved in the process of story-telling and autobiography.
Societal and Cultural Value: Records documenting the contemporary character, personality, intimacy, beliefs and spirituality of individuals within society.

In a 2010 essay Canadian literary archivist Catherine Hobbs states:

> To see more of the personal context of documentation, we need to understand as much as much as possible of the creator's intention and thoughts about documenting. Archivists should consider all psychological factors involved when individuals make/keep/destroy documents. The creator's relationship to documentation could involve many emotional and practical aspects…What were the choices to and motivation for creating the record?[62]

Assessing the impulses and intentions, appraisal decisions, and designations of value within the broader context of private records creation has for some time been the

[61] PIM Designations of Value are derived from the studies examined in this section of the thesis while Archival Designations of Value are based on the discussion of personal archives theory in chapter one. There are a number of similarities between PIM and archival designations of value. For example, "Identity Value" is similar to the conception of "Narrative Value" advocated by archivists Catherine Hobbs and Verne Harris in chapter one of this thesis. Comparatively, "Informational Value" relates to "Posterity and Legacy Value" while the archival designation of "Societal and Cultural Value" may encompass all of the PIM notions of value shown in this figure.

[62] Catherine Hobbs, "Reenvisioning the Personal: Reframing Traces of Individual Life" in *Currents of Archival Thinking*, Terry Eastwood and Heather MacNeil , eds. (Santa Barbara: Libraries Unlimited, 2010), pp. 227-228.

province of personal records archivists. Yet as Figure 2.2 reveals, personal information management research identifies many of the same values commonly attributed to records by personal archivists and serves to complement and reaffirm the already established value folksonomies of personal archives.[63] Given the precipitous rate at which digital material is created and accumulated, PIM research regularly examines how individuals decide what constitutes meaningful or insignificant items in the construction of their personal archives. This is of direct relevance to the provenance of personal archives and as such, should be leveraged and exploited by those seeking to preserve it in the interests of documentary heritage.

Personal Preservation of Digital Records

Following mediative processes of personal recordkeeping and private assessments of value, digital materials expand beyond their initial use and become much more than simple documentation. Here, individuals engage in initiatives to preserve their valued digital records for extended periods of time. Far from being a one-time static event, these preservation practices command an ongoing personal commitment to keep track of diverse digital records with unique curatorial needs displaced throughout multiple storage areas. Interestingly, the preservation of personal digital records is the one area of the pre-custodial environment memory institutions consciously seek to influence through awareness-raising campaigns and other outreach activities. For example, the Library of Congress website contains a series of WebPages dedicated to personal archiving where individuals may download a best-practice brochure titled “Preserving Your Digital Memories” and view a digital preservation awareness video which states “No matter what

[63]Value folksonomy is defined here as a classification of terminology often used by both individual records creators and archivists to describe the values of personal archives. Taxonomy is far too formal of a term to be used in discussions on the value of personal archives.

type of file you want to save, they all require the same essential preservation strategy: identify what you want to save; decide what is most important to you; organize the content; save copies in different places."[64] But is it really that simple and do best practices such as the ones proposed by the Library of Congress operate effectively in everyday digital life? The next section of this chapter investigates the realities of personal digital preservation through an analysis of pertinent PIM literature on the topic of localized and online digital preservation.

Localized Digital Preservation

In 2007, Microsoft researchers Gordon Bell and Jim Gemmell stated a six-hundred-dollar hard drive could hold one terabyte (one trillion bytes) of data which is "enough to store everything you read…all the music you purchase, eight hours of speech and 10 pictures a day for the next 60 years."[65] Comparatively, Neil Beagrie of the British Library contends that with current exponential increases in levels of computing power and storage capacity it will soon be possible "to envisage individuals being able to store the equivalent of all the texts in the Library of Congress on their PC."[66] All of these bits and bytes of personal data are housed within localized preservation environments consisting of the HDD of desktop and laptop PCs, USB drives, CDs, DVDs, but also

---

[64] Library of Congress, *Preserving Your Digital Memories and Why Digital Preservation is Important for You*, Available at http://www.digitalpreservation.gov/you/ (accessed 30 June 2011). Video transcript available at http://www.digitalpreservation.gov/videos/personal_archiving/index.html (accessed 30 June 2011).

[65] Gordon Bell and Jim Gemmell, "A Digital Life," *Scientific American* 296 (March 2007), pp. 58-65. As of 2011, a one terabyte external hard drive is sold by major Canadian electronics retailers for approximately one-hundred to one-hundred-fifty dollars while making external two terabyte hard drives available for a marginal increase in cost.

[66] Neil Beagrie, "Plenty of Room at the Bottom? Personal Digital Libraries and Collections," *D-Lib Magazine*, 11:6 (June 2005). Available at http://www.dlib.org/dlib/june05/beagrie/06beagrie.html (accessed 2 June 2011). Beagrie quotes Gordon Moore's seminal article on the increase of computing power "where there is roughly a doubling of the number of transistors on integrated circuits every 18 months for the same unit cost." See also Gordon E. Moore, "Cramming more components onto integrated circuits," *Electronics* 38:8 (April 1965), pp. 1-4. Available at download.intel.com/museum/Moores.../Gordon_Moore_1965_Article.pdf (accessed 2 June 2011).

within the memory media of external devices such as digital photo and video cameras, media players, smart phones, and more recently tablet technologies. Further, as digital information is inherently bound to the software necessary to interpret and render its binary sequences of zeros and ones, personal data is further distributed across multiple file formats many of which are proprietary in nature. Adding to the complexity of file formats is the issue of "lossy algorithms" where data is compressed to minimize file size but causes resolution to become distorted and pixilated with digital images and reduces certain frequencies in audio recordings. In short, these localized environments contain an ever expanding amount of digital files stored in myriad logical (software) and physical (hardware) carriers.

Conventional strategies for personal archiving in local environments involve backups (performed manually or automatically), exporting files to external storage media (CDs and DVDs), and the preservation of entire platforms (the computer, its peripherals, and installed software).[67] Periodic backups are the most basic form of data replication where redundant copies are created and stored to protect against loss caused by user errors, disk or other hardware failures, software errors, and natural disasters.[68] People may protect their file systems with either a full backup thereby copying the entire contents of the file system, or with an incremental backup which copies only those files that have been modified since the previous backup. As an archiving behavior, backing up also refers to the replication of data stored in media players, digital cameras, and

---

67 Catherine C. Marshall, "Rethinking Personal Digital Archiving, Part 1: Four Challenges from the Field," *D-Lib Magazine* (March/April 2008). Available at http://www.dlib.org/dlib/march08/marshall/03marshall-pt1.html (accessed 2 June 2011).

68 Ann L. Chervenak, Vivekanand Vellanki and Zachary Kurmas, "Protecting File Systems: A Survey of Backup Techniques," *Proceedings of the Joint IEEE and NASA Mass Storage Conference*, (March 1998). Available at http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.31.7765 (accessed 2 June 2011), pp. 17-31.

smartphones, which involves the synchronization of data distributed between these external devices and the system hosting the backup. A similar form of replication involves copying valuable digital items to external hard drives, CDs, DVDs, USB flash drives and other contemporary storage media, which is then labeled and placed in a physical storage area. As a long-term archiving strategy, this digital content may be migrated to successive physical carriers depending on the preservation regime followed by the individual. A final strategy is the full retention of entire computer platform (in place of exporting data and software), which may occur upon the purchase of a new and faster computer system.

Online Digital Preservation

While it may also be referred to as Cloud or Web 2.0 storage, all online digital preservation operates on the premise of a client-server relationship where an individual's digital records are stored on a server infrastructure they neither own nor control in regard to how often data is backed up or how long it is retained.[69] Quite often, individuals take advantage of free or moderately priced online storage and distribute the custody of their records across multiple online services (multiple servers) resulting in silos of digital storage as opposed to centralized repositories. Conventional strategies for online personal archiving include: the email-repository strategy; storing records on commercial file sharing platforms, social networking services, or a blog/podcast publishing service; and soliciting remote storage from online service providers.

[69] Personal computer users (the client) requests data from a more dynamic and often third party computer (the server) which responds by sending the requested data back to the user. Clients may request data from a variety of computer servers responding with database information, WebPages, email, or streaming media. Individuals also participate in this client-server relationship in social media when publishing blogs, updating Facebook status, tweeting, or posting images to Flickr or videos to YouTube.

With its almost unlimited storage capacity and proven utility as a personal information management tool, it seems logical that individuals would extend archiving practices to their email. Email has the ability to send and store many of the same file classes found in localized preservation environments through attachment or embedding options. Keeping these files online is in itself a preservation measure; however, individuals may choose to repatriate their data from commercial computer servers to bring it back under their control in localized storage in a process that may be referred to as "data liberation".[70] Email client software uses one of two protocols for message retrieval: Internet Message Access Protocol (IMAP) which stores messages on a mail server; and Post Office Protocol (POP) which stores messages on the HDD of the personal computer.[71] In order to capture and preserve an email account(s), a user must enable POP with their mail provider (Gmail or Hotmail), access their email account with POP client software (Thunderbird or Outlook), locate the email files (.msf or .pst) within the personal computer's file system, and preserve these files in a localized environment.

An extension of the email-repository strategy is to upload and store digital records with commercial file sharing platforms (YouTube, Google Docs, or Flickr), with a social networking service (Facebook, or Google+), or with a blog/podcast publishing service (Blogger, Twitter, or iTunes). Similar to email, individuals may choose to keep their digital files on Web 2.0 platforms or choose to re-capture this content for localized storage. Although some content uploaded to Web 2.0 applications may indeed be digital

[70] Data liberation is a relatively new phenomena brought on by the ubiquity of Google products. Internet-based groups such as The Data Liberation Front seek to inform the public on how people can export data from Google products such as Gmail, Picasa, Blogger, and YouTube if and when people decide to stop using these products. See http://www.dataliberation.org/home (accessed 30 July 2011).

[71] IMAP messages remain on a web server until the user deletes them from the inbox or other mail folders. When POP messages are retrieved from the mail server and stored locally they are typically deleted from the mail server. Google Mail, however, provides POP users with the option to keep messages on their servers even after they have stored them locally.

surrogates of files already kept in localized preservation, content such as "tweets" or Facebook posts irrevocably circumvent the HDD of the personal computer, while some content uploaded to YouTube, Flickr, or Blogger may be the only existing versions. Repatriating personal Websites and more advanced Web 2.0 content may be as simple as re-downloading files to a personal computer or clicking the export button, yet some content may require a more specialized software application such as Warrick, a utility that searches the Internet Archive, Google, Bing, and Yahoo for cached versions of WebPages and stores them on the user's personal computer.[72] In the case of social media, individuals may use utilities such as ArchiveFacebook, a web browser add-on which captures and exports all file data from their Facebook account to a directory on their personal computer thereby making it available for localized preservation.[73]

A final online preservation strategy is to deposit digital items with commercial cloud storage services:

> Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.[74]

---

[72] Frank McCown et.al, "Lazy Preservation: Reconstructing Websites by Crawling the Crawlers," *Proceedings of ACM WIDM* (2006), available at http://www.cs.odu.edu/~fmccown/pubs/lazyp-widm06.pdf (accessed 15 July 2011). See also http://warrick.cs.odu.edu/. As search engines crawl and index the Web, encountered pages are temporarily stored in a cache. Cached pages reduce network traffic and improve the responsiveness of the Web by providing clients with access to the search engine copy rather than the original server copy. Put differently, cached pages sit between the client request and the original server response.

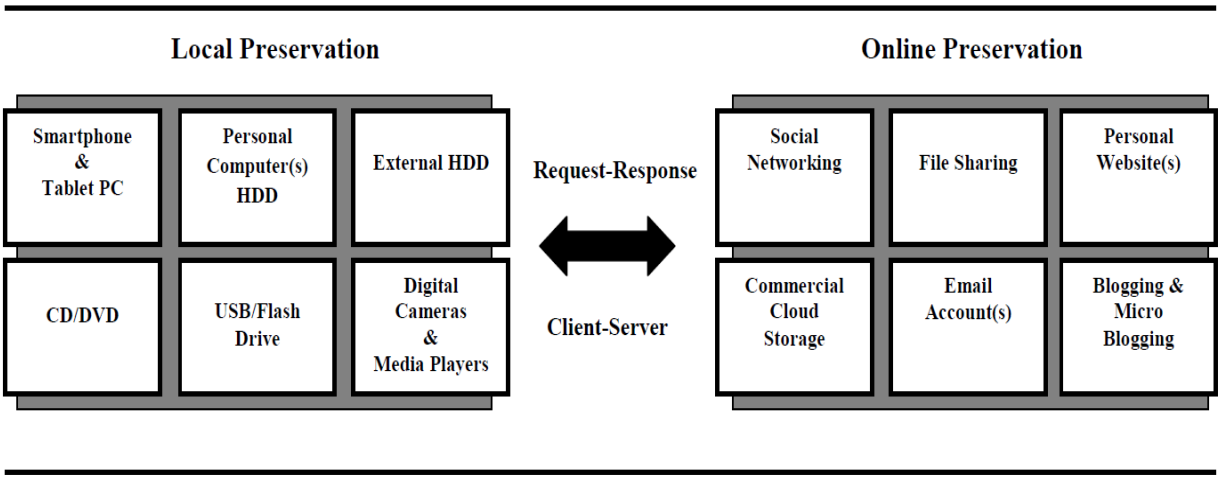
[73] Frank McCown and Michael L. Nelson, "What Happens when Facebook is Gone?" *Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries* (2009), available at http://portal.acm.org/citation.cfm?id=1555440 (accessed 15 July 2011). See also Mozilla Firefox add-ons page available at https://addons.mozilla.org/en-US/firefox/addon/archivefacebook/ (accessed 15 July 2011). This add-on captures and typically stores the data in C:\Documents and Settings\user\Application Data\Mozilla\Firefox\Profiles\. Similar services include TwapperKeeper, Twistory, and Keepstream.

[74] Peter Mell and Timothy Grance, "The NIST Definition of Cloud Computing (Draft)," National Institute of Standards and Technology, U.S. Department of Commerce, (January 2011), available at www.nist.gov/itl/cloud/upload/cloud-def-v15.pdf (accessed 15 July 2011).

Cloud storage is built on the client-server architecture where an individual transfers files from one personal computer to a commercial server where the files are stored and may be accessed from any personal computer with an Internet connection. This system is similar to using File Transfer Protocol (FTP) client software to transfer files over the Internet by uploading them to an FTP host server. Unlike the email-repository strategy, cloud storage provides for larger and greater numbers of files often at no financial cost or for a nominal storage fee. For example, Dropbox allows users to store 2GB of data on their servers for free but incrementally charges for amounts exceeding that limit.[75] Other cloud storage providers such as Mozy, ChronicleofLife, MyEvents, Legacy Locker, and Making Everlasting Memories provide ongoing storage and access, automatic and scheduled data backups, records of usernames and passwords, and encryption for all types and forms of personal files in exchange for subscription fees ranging from five-dollars per month to one-dollar per MB of permanent storage space. Repatriation of files stored within the cloud typically involves the subscriber re-transferring content back to their local environment or requesting a disk copy of their digital materials from the commercial service.

[75] See Dropbox, homepage, available at http://www.dropbox.com/ (accessed 15 July 2011). Dropbox currently offers 50GB, 100GB, and 350+GB of storage space. Dropbox also creates an exact mirror of a user's cloud content on their local PC or laptop HDD.

Figure 2.3 Personal Digital Preservation Environments



The degree to which personal digital preservation strategies, whether local or online, are successful is also subject to quotidian determinants both within and beyond the control of individuals. For instance, Microsoft researcher Catherine Marshall has found that although individuals often have the best intentions of adhering to preservation principles, everyday practices may belie these principles when people begin to implicitly rely on sporadic backups and unsystematic file replication as the primary modes of preservation.[76] Similarly, as digital preservation is an ongoing series of curatorial actions, individuals may not be able invest this required effort for all of their digital records all of the time, and some records may fall victim to what Marshall identifies as "benign neglect" brought on by the burden of digital stewardship.[77] Additionally, as the status and terms of service of email providers, social networking sites, file sharing platforms, and cloud storage providers changes over time, accounts and profiles may be deleted or deactivated by these entities without properly notifying those who have uploaded and

[76] Marshall, "Rethinking Personal Digital Archiving, Part 1".
[77] Ibid. Marshall describes benign neglect as setting aside digital records that matter in a special place and hoping they will be there when the time comes to retrieve them.

otherwise entrusted their data to them. Similar issues arise when companies discontinue the production of required peripheral computer hardware such as eight or five-and-one-quarter inch floppy disk drives, or when the array of proprietary digital formats becomes so opaque as to make individuals unsure of the consequences of their format choices.[78] A final element affecting the success of personal preservation strategies is the potential of malicious software (malware) attacks able to circumvent even the most trusted anti-virus software and system firewalls.

Summary of Chapter: Implications for Archivists

In the 1990s, archivists ruminated on the digital recordkeeping practices of the private sphere with some drawing parallels with corporate recordkeeping and others seeking to influence personal digital recordkeeping through pre-custodial intervention.[79] While both groups presented valid arguments, the latter approach was deemed impractical due to the temporal standpoint of archivists, while the former appeared oversimplified given that corporate and personal recordkeeping practices were not exactly alike. All the same these two themes remain relevant to the discussion of personal digital archiving. As will be discussed further in a moment, PIM research on personal computing behaviours and technologies is valuable because it straddles this intersection of the personal and the professional in ways that illuminate personal recordkeeping behaviour. And a more recent probing of the pre-custodial environment re-emphasizes its importance in archiving:

> [T]he archivist must also seek to understand the sequence in which the records may have been created, how the creator may have used the

[78] Marshall, "The Long Term Fate of Our Digital Belongings".
[79] See Richard J. Cox, "The Record in the Manuscript Collection," *Archives & Manuscripts* 24:1 (May 1996), p. 52; McKemmish, "Evidence of me…," pp. 28-45; and Cunningham, "Waiting for the Ghost Train," p. 58.

> records, and to what end(s). Doing so requires the study and analysis of the creator, not only in terms of *what* he or she was doing, but also *how* and *why*.[80]

Indeed, comprehending the intricacies of the pre-custodial environment has been of the utmost concern to archivists as it elucidates the provenance of personal records and is, most especially in the digital era, integral to their ongoing preservation. An issue, however, has been a stark inability to examine the nascent stages of personal records creation as they develop *in the here and now*. Moreover, the information about twenty-first-century pre-custodial environments that does exist is typically gleaned from seemingly analogous personal papers and creators who precede the digital era.[81] PIM research provides a lens through which archivists may view not only contemporary recordkeeping practices of individuals but also private appraisal decisions and personal digital preservation strategies closer to the point of their actual performance.

Some of the personal recordkeeping practices discussed in this chapter have been derived from PIM studies taking place in work environments as opposed to strictly home settings, and yet these should not simply be conflated with institutional records and archival management practices for a number of reasons.[82] First, discussions of government and corporate digital archives often centre on appraisal criteria and systematized records classification according to high-level functions, activities, and business processes with little consideration given to how these methods and directives are

---

[80] Jennifer Meehan, "Rethinking Original Order and Personal Records," *Archivaria* 70 (Fall 2010), pp. 38-39. Emphasis in original.

[81] Ibid., pp. 40-43. For example, in her excellent discussion of the original order of personal records, Meehan references the archival collections of three individuals whose lives, and therefore records creation, ended before many digital documentary forms discussed in this chapter even existed.

[82] The workplace studies include: Henderson, "Personal Document Management Strategies" (knowledge workers); Barreau and Nardi, "Finding and Reminding" (managers); Mackay, "More than Just a Communication System" (full-time researchers, managers, computer scientists, academic professionals, and administrators); and Whitaker and Sidner, "Email overload" (office workers).

applied and followed by actual human records creators on a day-to-day basis.[83] PIM studies on the other hand, focus on the individual and how he or she interacts with their computing environment with little if any discussion of business functions or directives and standards for corporate recordkeeping. Although PIM research may at times be carried out in formal office settings, it privileges the discovery of day-to-day personal computing behaviours in those office settings over the analysis of the functions those offices perform in an organization. And, as archivist Catherine Hobbs observes, there is an interplay of the personal and the professional in private archives where it is possible to identify how private life influences one's work.[84] Inverting this observation, activities learned and performed in occupational life, such as choices of hardware and software or strategies for email and file management, may inform how digital records are created, managed, and preserved in private life. While the precise mechanics of personal recordkeeping and archiving at home and at work may not be exactly alike, there will undoubtedly be points of convergence and overlap as, with the growing predominance of mobile devices and cloud computing, the boundaries of these two realms may at times become "very hard to delineate clearly, if at all."[85]

The personal recordkeeping strategies identified in a number of PIM studies reviewed in this chapter may be incorporated into archival workflows in the areas of intellectual and physical arrangement of personal archives as this information may come to guide archivists through, for instance, the reconstruction of original order. This

---

[83] See Inge Alberts et al., "Bridging Functions and Process for Records Management," *Canadian Journal of Information and Library Science*, 34:4 (December 2010), pp. 365-690.
[84] Hobbs, "Reenvisioning the Personal," pp. 223-224. To illustrate this point, Hobbs refers to the personal archives of Canadian writer Carol Shields. Hobbs was able to pick up on the interplay and connection between events occurring in the daily life of Shields and her work as a novelist.
[85] Leighton John et. al., "Digital Lives," p. 32.

reconstruction of original order may mean replicating the folder directories of a *frequent-filer* or *pro-organizer* as they appear at the point of acquisition. Conversely, reconstructing the seemingly chaotic, but nevertheless meaningful, original order of records belonging to a *no-filer* who relied on keyword searching for the rediscovery of their files may require that the archivist ascertain how certain desktop search engines indexed files and what filtering options were available to retrieve data from personal computer hard drives.[86] Maintaining some semblance of original order is also important for users of personal digital archives as it provides cues on discovering the relationships between records that would otherwise be lost if records were haphazardly reorganized by archivists. While archivists many not be able to analyze personal recordkeeping behaviours in the same way as they do corporate records management, they may come to develop frameworks for analyzing patterns of those behaviours based on personal information management research and gain better pre-custodial understandings of the meta-cultures and practices of personal recordkeeping. Additionally, appraisal decisions disclosed by PIM research participants may influence the way archivists conduct their appraisals of personal digital archives. Criteria such as identity, personal memory, familial and historical, emotional and sentimental, as well as posterity and legacy value should serve as benchmarks in the process of selecting records within a personal digital archives for permanent retention.

PIM research is also relevant to archival work in discovering the true breadth of personal digital fonds or collections. Never before have the records that constitute a personal fonds or collection been so widely distributed across multiple locations, and as

[86] Before a keyword search can be executed, an index database (list of files and their locations) must be built. While this index building often occurs automatically, certain computer drives must be manually indexed by the user.

shown in this chapter, PIM research elicits key information on how and where individuals preserve their valuable records in both on and offline environments. This is especially germane in an era where personal digital records are being kept by non-familial third parties more for their commercial value than for their cultural and historical value. Indeed, archivists must know where to look for personal digital records before they can appraise and select them in the interest of producing a truly representative personal fonds or collection. The recordkeeping behaviours, designations of value, and digital archiving strategies discussed in this chapter may also figure prominently in archival descriptions as this information reveals how individuals create, manage, and use their digital records within certain contexts. While it is true that PIM studies do not address the issue of locating those individuals within society whose records warrant archival preservation, these studies do identify why, how, and where individuals archive their digital records in certain temporal and technological contexts and as such, should inform the mediation of personal digital records in both the pre-custodial and archival environments discussed at the beginning of this chapter.

The following chapter explores how archival institutions are meeting the challenges of personal digital records identified in the first and second chapters of this thesis. The third chapter examines two approaches to personal digital archives: rescuing electronic data from obsolescence; and front-end digital acquisition and preservation methodologies. It concludes with a discussion of embryonic strategies for the future of personal digital archives.

CHAPTER THREE

RECENT INNOVATIONS AND RESEARCH IN ARCHIVING PERSONAL DIGITAL RECORDS

> While it is encouraging to find so many institutions see the value of collecting born-digital materials, it is disconcerting to consider the state these records must be in if most institutions cannot even estimate how much of them are present in their collections, have given no thought to what kinds of records they acquire, and have not developed policies addressing how to manage and preserve them overtime.[1]

Having discussed traditional archival and personal information management perspectives on personal archives, the third chapter of this thesis reviews and analyzes exemplar approaches to archiving personal digital records in contemporary memory institutions. Some of the approaches examined in this chapter aim to acquire and recover data from personal electronic records of the recent past (fifteen to twenty years ago). Some seek only to process the digital records of the relative present. Others strive to discover innovative approaches for the future of personal digital archiving. In short, there is no one, all encompassing approach that may be used for personal records in the digital age as archivists responsible for these materials must triage collections comprised of data dependent on obsolete and fragile media, contend with personal records relative to their generation, while at the same time strategizing for future acquisitions and the ongoing preservation of digital collections already in their care. Given this division of archival labour, chapter three examines contemporary archival approaches to personal digital records in three sections: rescuing electronic data from obsolescence; front-end acquisition and preservation methodologies; and embryonic strategies for the future of

[1] Ben Goldman, "Bridging the Gap: Taking Practical Steps Toward Managing Born-Digital Collections in Manuscript Repositories," *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage* 12:1 (Spring 2011), p. 13.

personal digital archives. The chapter opens with a review of the seminal work done on the Salman Rushdie digital archives at Emory University and the Michael Joyce fonds at the Harry Ransom Center at the University of Austin. It then assesses the Library and Archives Canada, Trusted Digital Repository (LAC-TDR), the Personal Archives Accessible in Digital Media (Paradigm) project in Britain, and the Canadian Artefactual Systems Inc. Archivematica software. It concludes with an evaluation of an innovative front-end approach advocated by the Digital Lives Research project at the British Library, a discussion of a prospective technology to be used by personal records creators, and a survey of new approaches to the creation of archival metadata for personal records.

Rescuing Electronic Data from Obsolescence

In 2002, archivist Karyn Taylor surveyed a number of archival repositories to determine the status of personal electronic records in terms of acquisition and preservation policies. Understandably, the responses from national, provincial and university institutions differed but as Taylor notes "there is no consensus on how institutions deal with personal electronic records. Furthermore, several of the institutions appear to be just beginning to give the matter thought."[2] In a 2008 survey of one-hundred-twenty-five collecting repositories within the United States, archivist Susan E. Davis discovered "that archivists are incorporating born-digital records and papers into their collections without necessarily altering existing policies to do so."[3] In a 2010 survey of small and medium sized archival institutions within Canada, the InterPARES III

[2] Karyn Taylor, "From Paper to Cyberspace: Changing Communication Technologies and the Implications for Personal Records Archivists," University of Manitoba, Department of History (Archival Studies) MA Thesis (2002), p. 89. Taylor's survey data was drawn from respondents employed by: the University of Manitoba Archives & Special Collections; the Provincial Archives of Manitoba; the National Archives of Canada; the National Archives of Australia; the S.J. McKee Archives, Brandon University; and the National Library of Australia.

[3] Susan E. Davis, "Electronic Records Planning in 'Collecting Repositories," *The American Archivist* 71 (Spring/Summer 2008), p. 185.

project learned that while provincial, university, community, and library-level repositories currently accepted digital records, it became evident "that many institutions do not have adequate policies or procedures in place to deal with the acquisition and/or preservation of digital records."[4] In short, many archival institutions are consciously aware of the value of personal digital archives, but have yet to obtain the means to acquire and preserve them. As a result of this strategic deficit, archival institutions continue to acquire what may be referred to as ancestral computer platforms, software, and storage media for which processes of digital archeology must be performed in order to rescue electronic data from unstable technologies and certain erasure.[5] Alternately, the foregoing of these measures will in all likelihood mean the difference between a personal digital archive truly representative of an individual and their activities, and a personal digital archive containing chronological inconsistencies and other stark discontinuities in documentation. To explain and offer potential remedies to problems associated with rescuing obsolete personal electronic records, this section now turns to a review of two pioneering archival case studies.

In 2006, the Manuscript, Archives, and Rare Book Library (MARBL) at Emory University acquired the personal archives of celebrated writer Salman Rushdie, which in addition to approximately one-thousand linear feet of journals, correspondence,

---

[4] InterPARES III Project, "The Acquisition of Digital Records at Small and Medium Sized Archives in Canada: Survey Research from InterPARES 3" available at http://www.interpares.org/display_file.cfm?doc=ip3_canada_dissemination_post_van-dijk~malmas_slais-research-day-2011.pdf (accessed 25 August 2011).Small and medium sized archival institutions include: museums and heritage centres, territorial and provincial archives, university and college archives/special collections, community archives and libraries, as well as religious and municipal archives.

[5] One of the earliest uses of the term "ancestral computing" is found in a paper delivered by Jeremy Leighton John titled "Adapting Existing Technologies for Digitally Archiving Personal Lives: Digital Forensics, Ancestral Computing, and Evolutionary Perspectives and Tools," *iPRES 2008: The Fifth International Conference on Preservation of Digital Objects*, British Library Conference Centre, St. Pancras, London, September 29-30, 2008.Available athttp://www.bl.uk/ipres2008/presentations_day1/09_John.pdf (accessed 25 September 2011).

photographs, manuscripts, and audio-visual material, contained a substantial component of digital materials. The Rushdie acquisition included a number of outdated Macintosh computer platforms such as a Performa 5400, three iterations of the Macintosh PowerBook, and a SmartDisk portable hard drive containing born-digital materials such as manuscripts, email correspondence, daily calendars, diaries, as well as downloaded and cached content from the Internet culminating in approximately forty thousand files or 18 gigabytes of data.[6] Given the complex hybrid nature of the Rushdie archives, Emory was required to take an innovative approach to traditional archival processing and created the Born-Digital Archives working group comprised of archivists as well as digital systems experts.

The early steps of processing the Rushdie digital archive involved an assessment of the health of the hardware, the performance of system diagnostics, as well as the identification of file formats and encryption. Following this, data duplication via disk imaging for all four obsolete personal computers was performed to capture original file and system data. By using disk imaging Emory was able to create exact bit-by-bit copies of disk drives to authentically replicate the content, structure, and functionality of the original personal computers. This procedure was assisted by a suite of digital forensic tools including the Duke Data Accessioner, an extremely user-friendly open source program designed to transfer data from physical media or computer directories to a dedicated file server or local storage directory for basic preservation, appraisal,

---

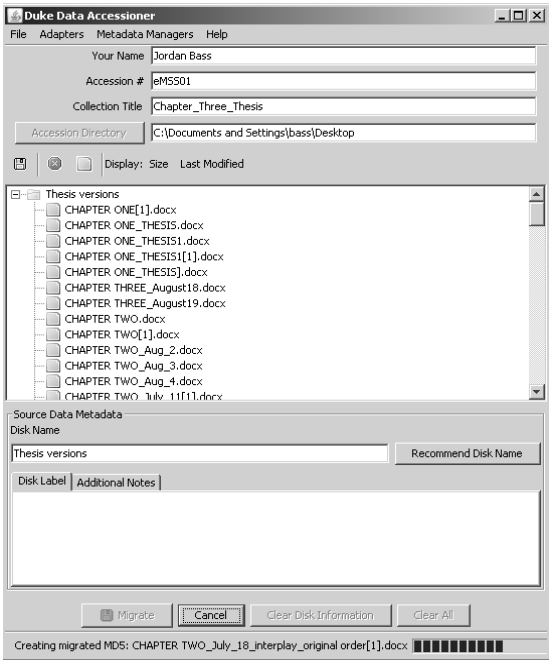
[6] "Salman Rushdie Papers, 1947-2008," Emory University, Manuscript, Archives, and Rare Book Library, http://web.library.emory.edu/blog/salman-rushdie-papers-finding-aid-now-available-online (accessed 11 August 2011); Laura L. Carroll, "Arranging and Describing Born-digital Archives: The Salman Rushdie Papers at Emory University," PowerPoint Presentation, *Society of American Archivists Conference*, Chicago, Illinois, August 26, 2011, available at http://e-records.chrisprom.com/wp-content/uploads/2011/08/PAERWhatWorksNow.pdf (accessed 25 August 2011); and Mary J. Lofus, "The Author's Desktop," *Emory Magazine* (Winter 2010) available at http://www.emory.edu/EMORY_MAGAZINE/2010/winter/authors.html (accessed 16 August 2011).

arrangement and description.[7] Once the entire extent of the Mac platforms was extracted and copied to a stable offline processing environment, the working group was able to review and index directories as well as file types and associated applications (such as MacWrite Pro), issue SHA1 and MD5 checksum hash values, generate persistent identifiers, and record this and other value-added metadata in spreadsheets.[8]

[7] Laura L. Carroll, "The Salman Rushdie Papers at Emory University: Processing a Born Digital Manuscript Collection," PowerPoint Presentation, *Association of Canadian Archivists Institute on Personal Archives*, Toronto, Ontario, October 29 and 30, 2010.

[8] Value-added metadata refers to file names, directory paths, creation and modification dates that were recorded in addition to verdicts on whether a file would be released to the public "as is," "redacted," "restricted," or "emulation only." Series and sub-series information was also assigned to each file. See Erika Farr, "When Papers Aren't Just Paper: Hybrid Archives at Emory," PowerPoint Presentation, *Digital Lives Conference*, London, England, February 9, 2009; and Laura L. Carroll Erika Farr, "Hybrid teams for hybrid archives: Collaboration and born-digital archives," PowerPoint Presentation, *Association of College and Research Libraries: Rare Books and Manuscripts Section (RBMS) Conference*, Philadelphia, Pennsylvania, June 22-25, 2010. Executing a checksum is analogous to taking a fingerprint of a digital object. For example, the Message-Digest (MD5) checksum for this chapter is 1fe0cd3920053d4eabc1d6e555e33b77 Repeating this same process later should generate the same hash value. If this Word file has been modified, the value generated will be different.

Figure 3.1 The Duke Data Accessioner User Interface[9]



Following the duplication and manipulation of the Rushdie computers, the Emory team created a standalone, off-line workstation at the MARBL reading room equipped with a database containing hundreds of available files from Rushdie's Performa 5400. This database includes keyword search and browse functionality as well as full-text access to materials such as drafts of manuscripts, email messages, faxes, and other digital documents. Most notable, however, is the emulated environment of the Performa 5400, an exact replica of its operating system, desktop directory structure, and software applications which Emory created using an open source emulator program called

[9] Duke University, Duke Data Accessioner, Version 0.3.0. (Stable). Figure 3.1 shows the Data Accessioner migrating data from a file directory titled "Thesis Versions" on a portable USB drive to the desktop of a personal computer. MD5 checksums are also generated for each file while a PREMIS-based xml manifest is also created to document file identification and metadata extraction activities. Permission for Duke Data Accessioner screen shot obtained from Seth Shaw, Electronic Records Archivist, University Archives, Duke University Libraries.

SheepShaver.[10] In this emulated environment, researchers are able to explore folders and sub-folders as they were arranged by Rushdie, manipulate the Performa desktop, and view drafts and notes rendered in their original format to gain a better understanding of Salman Rushdie's literary production and computing behaviours.[11] To summarize, Emory recovered at-risk data from obsolete personal computers, duplicated this captured data for preservation and workable copies, and provided twofold access to the Rushdie digital archive through a searchable database and an emulated environment to present its content, context, and structure as authentically as possible.

A comparable case study took place at the Harry Ransom Centre (HRC) at the University of Austin with the fonds of hypertext author Michael Joyce. In his seminal electronic publication titled *Afternoon, a story*, Joyce incorporated a software program from the early 1990s called Storyspace to create one of the first fictional works relying on hypertext functionality to progress through the story.[12] Acquired in 2005, the Michael Joyce Papers are comprised of sixty manuscript boxes of paper-based materials in addition to three-hundred-seventy-one, 3.5 inch floppy disks totaling 211 kilobytes of data written by programs on Macintosh "classic era" personal computers.[13] Much like the Rushdie archives acquired by Emory, the age and instability of electronic media on which the data was stored, in this case floppy disks, required the HRC access and migrate the Joyce digital content to a more contemporary and stable computing environment. To this end, ancestral Macintosh hardware with integrated floppy drives were used to access

[10]See SheepShaver homepage at http://sheepshaver.cebix.net/ (accessed 25 August 2011).
[11] Carroll, "The Salman Rushdie Papers." See also, "Rushdie Researcher Workstation Tutorial." Video available at http://www.youtube.com/user/emorylibraries (accessed 09 August 2011).
[12] A web edition of *Afternoon* is available from Norton Anthology of Postmodern American Fiction at http://www.wwnorton.com/college/english/pmaf/hypertext/aft/ (accessed 21 September 2011).
[13] Catherine Stollar and Thomas Kiehne, "Guarding the Guards: Archiving the Electronic Records of Hypertext Author Michael Joyce," *New Skills for a Digital Era*, Washington, D.C. May 31-June 2, 2006. Available at www.lib.az.us/diggovt/documents/pdf/4_Stollar_Kiehne.pdf(accessed 10 August 2011).

content on each disk, which was then copied to directories on a Windows-based workstation where virus checks were performed. In the event of disc or file errors, archivists at HRC adopted an outdated commercial program, Norton Utilities, to recover files that otherwise could not have been copied to the stable workstation environment.[14]

The archivists at the HRC recorded all initial and subsequent processing information in an Excel Spreadsheet and created a thorough item-level inventory of copied file content with metadata such as file name, size, and format as well as dates of creation and modification. MD5 hash values were also generated for each individual file via batch processing which, in addition to establishing integrity control, supported the HRC archivists in provenance auditing and checks for file redundancy during surveying and appraisal. Following the creation of working copy masters to avoid repeated processes of file extraction and to preserve the bitstream (original binary code) recorded on floppy storage, electronic files were then intellectually arranged within archival series and deposited to an institutional repository built on a DSpace system architecture designed to preserve and provide access to many different types of digital content.[15] Access to the digital component of the Michael Joyce Papers, much like the digital archives of Salman Rushdie, is not made available via the World Wide Web but rather through a dedicated laptop workstation housed within a reading room at the HRC where researchers may freely download electronic files and their associated metadata from the DSpace repository.[16]

---

[14] Thomas Kiehne, Vivian Spoliansky, and Catherine Stollar, "From Floppies to Repository: A Transition of Bits, A Case Study in Preserving the Michael Joyce Digital Papers at the Harry Ransom Center," Unpublished paper, May 2005, available at https://pacer.ischool.utexas.edu/handle/2081/941 (accessed 25 September 2011).

[15] For more detailed information about DSpace see http://www.dspace.org/ (accessed 25 September 2011).

[16] Matthew G. Kirschenbaum, *Mechanisms: New Media and the Forensic Imagination*, (London, England: The MIT Press, 2008), pp. 207-211. In his discussion of researching the Joyce digital papers Kirschenbaum

Assessment of the Emory Library and Harry Ransom Centre Approach

The Salman Rushdie and Michael Joyce digital archives stand as two superlative examples of how archivists may rescue personal digital records from imminent deterioration through processes of digital archeology. For example, as many of the digital records within the Joyce archives were acquired some fifteen to twenty years after their initial creation, archivists spent a significant amount of time and resources on data-recovery from fragile and decaying storage media, most notably in researching workaround procedures for unreadable disks due to errors brought on by hardware and software incompatibilities.[17] Indeed, as HRC archivist Catherine Stollar Peters notes, "no utility existed that would perform all the digital archeology tasks we desired at one time."[18] Similarly, Emory commissioned the skills of software engineers to extract the data from Rushdie's computer hard drives and to emulate his Performa 5400.[19] The degree to which digital archeological methods such as these are required is correlated with the interval of time between initial data creation and archival acquisition. The validity of this statement is also supported by online postings from individuals who have been "through floppy hell" to recover their data trapped on obsolete 5.25 and 3.5 inch disk media as most current computer operating systems do not support floppy drive peripherals and controllers.[20] Fortunately, computer enthusiasts are beginning to develop workarounds for these technological obstacles such as Device Side Data's FC5025 USB

---

notes that he was not permitted to make copies of any electronic materials and that any screenshots made were printed out and paid for just like Xerox copies.

[17] Catherine Stollar Peters, "When Not All Papers Are Paper: A Case Study in Digital Archivy," *Provenance* 24 (2006), p. 27.

[18] Ibid.

[19] Lofus. In addition to archivists, the Born Digital Archives working group at Emory consisted of a senior engineer formally of IBM and a software engineer formally of Georgia State University.

[20] Gadgetmind (23 January 2011). 5.25" / 5 1/4" floppy drive-USB?Msg 7. Message posted to *Dos Game Discussions* available at http://www.dosgames.com/forum/about13481.html (accessed 25 September 2011).

5.25 inch floppy controller which allows users to access floppy disk media through USB 2.0 or 1.1 connection ports standard on current desktop and laptop computers.[21]

Both the Emory and HRC case studies employed migration as their primary preservation strategy, which is a technique defined as processes of replication occurring "between instances of the same type of storage medium, from one medium to another, and from one format to another."[22] In the Emory example, the migration of an entire Mac platform was followed by the reconstruction of its original functionality, directory structures, and GUI representation via emulation. In contrast, significant digital portions of the Michael Joyce Papers, most notably the Storyspace hypertext authoring and reading software, could not be emulated due to "copyright concerns, continued distribution of Storyspace by Eastgate Systems, and a lack of programming staff and time."[23] So while the HRC recognizes that researchers may wish to examine digital records in their original formats and representations, it also acquiesces to the bottom line that specialized staff with knowledge of both ancestral and contemporary hardware, software, file formats, programming languages, and databases are integral to making emulation, of the Rushdie digital archives variety, a feasible option for access.[24] The provision of access to personal digital archives at both Emory and the HRC is also of particular note as in both cases it is analogous to practices based on traditional paper-

---

[21] See http://www.deviceside.com/ (accessed 25 September 2011). The process of recovering data using the Device Side Data floppy controller stored is documented in a post at http://www.spellboundblog.com/ (accessed 25 September 2011) where the blogger provides a detailed description of rescuing data from 5.25 inch floppy disks dating back to 1984. A comparable floppy controller called Catweasel provides support for disks formatted by Amiga, Atari, Commodore, in addition to PC and Macintosh.

[22] David S.H. Rosenthal et al., "Requirements for Digital Preservation Systems: A Bottom-Up Approach," *D-Lib Magazine* 11:11 (November 2005). Available at http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html (accessed 25 September 2011).

[23] Peters, p. 33. Peters also notes that "Storyspace only runs on Windows or Macintosh operating systems, but the same program does not run on both nor does a file written in Storyspace 1.5 run properly in Storyspace 2.5."

[24] Ibid., p. 34.

based personal collections wherein researchers must physically visit the repository to interact with the records at the emulation and DSpace repository workstations. This suggests that while the transitory nature of born-digital or digitized records lends itself to online access, such is not always the case with personal digital archives.

In conclusion, given that digital archeology, migration, and emulation methods for personal digital records are still in their infancy stages, it may not yet be practical or economically feasible for archival intuitions to dedicate the amount of time and resources required to process multiple personal digital collections using the methodologies formulated at Emory and the HRC. Yet, as institutions continue to acquire personal archives at the end or near end phases of records creation, these types of retroactive procedures seem unavoidable. It seems prudent, then, that archivists factor this into their collection development, acquisition, and preservation policies as the recovery of data from a Macintosh Performa and 3.5 inch floppy disks will conceivably be succeeded by the recovery of data from multiple generations of Apple and Microsoft technologies and countless versions of portable storage media. Moreover, given the ubiquity of personal computers, archival institutions may anticipate performing data recovery, migration, and emulation for records created not only by literary figures such as Rushdie and Joyce, but also for those records created by individuals from many different professions and walks of life.

Front-End Acquisition and Preservation Methodologies

> Future work with personal archives can be expected to be increasingly proactive and entail a close understanding with and involvement of originators and their families and friends. The single most important consequence of the increasingly digital nature of personal archives is the

> need to preempt inadvertent loss of information by providing advice and assistance.[25]

The second section of this chapter examines current approaches to archiving personal digital records that may be employed in the present to ameliorate the degree to which archival institutions expend their resources processing materials dependent on obsolete technologies in the future. As content reviewed in this chapter adheres or makes reference to the Open Archival Information System (OAIS) model, a brief description of OAIS elements precedes an examination of three digital initiatives. They are front-end approaches to personal digital archives that are best suited for those creators and their records already identified by way of proactive documentation strategy or other macro-level appraisal models.

The Open Archival Information System (OAIS) Model[26]

An OAIS-based digital archive consists of "an organization of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community."[27] Originally developed by the Consultative Committee for Space Data Systems (CCSDS) for the long-term preservation of digital information generated by space agencies, the OAIS reference model has since become a fundamental component of digital initiatives in many archival institutions and as of 2003 is an approved ISO standard.[28] The OAIS model specifies how digital records should be

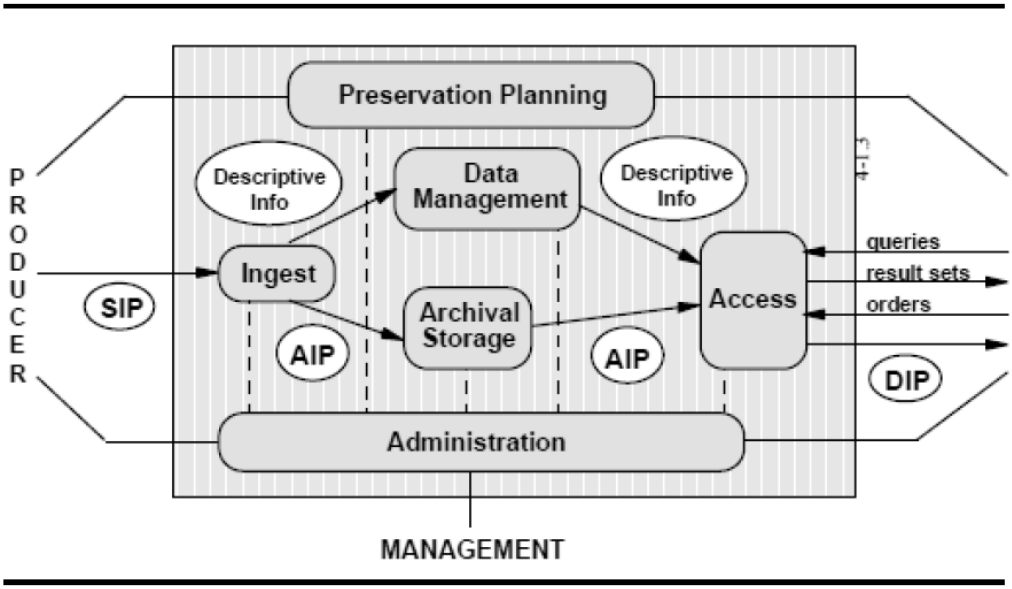
---

[25] John, "Adapting Existing Technologies for Digitally Archiving Personal Lives".
[26] See Figure 3.2 CCSDS - OAIS Functional Entities.
[27] Consultative Committee for Space Data Systems, *Reference Model for an Open Archival Information System (OAIS) Blue Book*, (January 2002), p. 1-1.
[28]Christopher A. Lee, "Defining Digital Preservation Work: A Case Study of the Development of the Reference Model for an Open Archival Information System," University of Michigan School of Information, PhD Dissertation, (2005), pp. xxvi-xxvii. See also International Organization for Standardization, "ISO 14721:2003, Space Data and Information Transfer Systems - Open Archival Information System – Reference Model," available at http://www.iso.org/iso/catalogue_detail.htm?csnumber=24683 (accessed 09 August 2011).

preserved from the point when they are received into a digital storage environment as a *submission information package* (SIP), through subsequent preservation management as an *archival information package* (AIP), to the creation of a *dissemination information package* (DIP) for a community of users. These three information packages, or conceptual containers of digital objects and their associated metadata, correlate with three external entities: personal or organizational records creators (*producer*), a designated community of users (*consumer*), and persons or organizations assuming management and policy responsibilities for the OAIS (*management*). As a reference model, the OAIS "has been widely accepted by the digital preservation community as a key standard for digital repositories" and aims to be as context-neutral as possible.[29] However, it is important to note that the OAIS model is a "conceptual reference architecture and does not specify a concrete implementation" and is often interpreted and incorporated into archival workflows in very different ways.[30] The following section of this chapter examines how digital records are processed through the OAIS stages of pre- and post-ingest.

[29] Susan Thomas, et al. *Paradigm: Workbook on Personal Digital Archives*, (Oxford: Bodleian Library, 2007), p. 3.
[30] Wolfgang Wilkes et al.. "Towards Support for Long-Term Digital Preservation in Product Lifecycle Management," *The International Journal of Digital Curation* 1:6 (2011), p. 283.

Figure 3.2 CCSDS - OAIS Functional Entities[31]



Library and Archives Canada Trusted Digital Repository (LAC TDR)

> In all domains the amount of digital information is increasing at a rapid rate, this raises crucial questions of preservation. Our intellectual capital, as laid down in educational, scientific, public, cultural and other intellectual resources, is increasingly at risk by the volatile character of digital objects and the rapid developments in information technology.[32]

In 2004, the Library and Archives Canada (LAC) embarked on a multi-year development project to establish a Trusted Digital Repository (TDR) consisting of policies, procedures, and a suite of technical tools and services to meet the challenges posed by the

[31] Ibid., p. 4-1. This graphical representation of the OAIS functional entities is taken from the CCSDS Blue Book. This exact image is often used in archival and digital preservation literature discussing the OAIS reference model and digital preservation in general.
[32] Library and Archives Canada, "Digital Initiatives at LAC, Digital Projects, LAC Trusted Digital Repository," available at http://www.collectionscanada.gc.ca/digital-initiatives/012018-4000.01-e.html (accessed 09 August 2011).

acquisition, management, and preservation of documentary heritage in the digital era.[33] Though planned primarily for the purposes of preserving the electronic records of the Government of Canada and the electronic publications of Legal Deposit, LAC may yet extend TDR workflows to incorporate personal digital records given the institution's long time commitment to the acquisition of documentary heritage materials generated by non-governmental entities. At present there are fourteen areas of LAC acquiring private records which are organized by theme (Social Archives or Literary Archives) or by media type (Cartographic and Geomatic Archives or Philatelic Collections).[34] Recently, archivists from these fourteen areas reviewed the status of the collections they oversee and have identified digital records as a key element of "emerging" private archives.[35] Although these archivists believe digital records do not fundamentally change the social activities they seek to document, they recognize that these materials represent a shift in what they are trying to acquire from analog materials to born-digital photography, manuscripts, and architectural records, for example.[36]

Given the national significance of the private archives held at LAC, it is important to examine how this documentary heritage may be processed within the institution's TDR based on how LAC is planning to process its electronic government records. Although

---

[33] Pam Armstrong, "Becoming a Trusted Digital Repository, Library and Archives Canada: Managing Information in the Public Sector, Meeting the Challenge," *Ontario Information Management Conference*, Toronto, Ontario, April 28, 2009, available at www.verney.ca/opsim2009/presentations/783.pdf (accessed 09 August 2011).

[34]Library and Archives Canada, Acquisition Priorities: Private Archives Working Group, "Final Report: Acquisition Priorities: Setting Priorities for Private Archives," The fourteen private archives sectors at LAC are: Art, Photography, Portraits, Film and Broadcasting, Literature, Performing Arts, Music, Philately, Politics, Governance, Social Archives, Economic Archives, Cartography and Geomatics, and Architecture. Available at http://www.collectionscanada.gc.ca/modernization/012004-2055.01-e.html#anc7 (accessed 15 August 2011).

[35]Ibid.

[36]Ibid.

the LAC TDR is not yet fully operational, the following section describes how LAC plans to implement and develop its massive digital archiving project.[37]

Pre-Ingest Activities

Within the context of the LAC TDR, an electronic record is defined "as both the digital object and its associated metadata" whereby the digital object (such as a relatively simple word processing file or more complex web publication files) is "inviolately linked to its complete metadata, including metadata generated by the record creator before transfer to the archives as well as post transfer metadata documenting LAC custody of the record."[38] Creator generated metadata is integral to LAC TDR ingest workflows as this data holds important information on the context of the digital object's creation, is key to the discovery of relationships between aggregated digital objects, and provides other facets of pre-ingest information that would otherwise be extremely difficult to capture retroactively. For example, creator generated metadata is essential to provide a reasonable basis for the presumption of the authenticity of electronic records and is used to verify the reliability and integrity of those records upon ingest and through ongoing maintenance within archival custody.[39] In short, metadata "is the only way to represent the context in which the record is created, when, by whom, for what purpose or to achieve what goal."[40]

---

[37]Library and Archives Canada, *Departmental Performance Report 2010-2011*. Available at http://publications.gc.ca/collections/collection_2011/collectionscanada/SB1-4-2011-eng.pdf (accessed 13 December 2011). As of December 2011, the LAC TDR is still in the early stages of implementation. LAC notes that building a TDR "proved to be more challenging than expected."

[38] Greg Bak and Pam Armstrong, "Points of convergence: seamless long-term access to digital publications and archival records at Library and Archives Canada," *Archival Science* 8:4 (2008), pp. 284-286.

[39] Ibid., p. 287. This metadata requirements for LAC TDR ingest are informed by the findings of the InterPARES Project :I Authenticity Taskforce, the ISO 15489:2001(records management), and the ISO 23081:2006 (managing metadata for records).

[40] Robert Nahuet, "The Management of Textual Digital Archives: A Canadian Perspective, Library and Archives Canada and Federal Government Institutions," *Atlanti*, 17:1-2 (2007), p. 42. Available at www.iias-trieste-maribor.eu/fileadmin/atti/2007/Nahuet.pdf (accessed 24 August 2011).

Electronic records may be submitted to the LAC TDR through ingest channels such as File Transfer Protocol (FTP) and Simple Object Access Protocol (SOAP), or retrieved by ingest agents such as the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) from OAI-compliant repositories.[41] At present, the LAC TDR plans to ingest records that are derived exclusively from government departments according to formalized agreements known as Records Disposition Authorities (RDA), and more specifically from Electronic Documents Records Management Systems (EDRMS) such as the Government of Canada RDIMS, an enterprise-wide software application and networked platform used by participating government departments and agencies in the creation, management, and disposition of electronic records.[42] However, as electronic records and metadata created by one department's RDIMS instance are markedly different from that of another, LAC has established a number of pre-ingest tools and standards to regulate records creation and to facilitate the transfer of RDIMS content to the TDR. For example, LAC has implemented a Local Digital Format Registry (LDFR), a Records Management Metadata Standard (RMMS), a related Government of Canada, Records Management Application Profile (GC RMAP) and an eRecord Transfer Application (eRTA) to ensure records destined for TDR ingest will arrive in a file format suitable for long-term digital preservation and persistent access.[43]

---

[41] Ibid., p. 284.

[42] The Government of Canada Records Document Information Management System (RDIMS) was selected in 1996 for electronic records management. RDIMS is managed by Public Works and Government Services Canada. See Johanna Smith and Pam Armstrong, "Preserving the Digital Memory of the Government of Canada: Influence and Collaboration with Records Creators," Mission Day Conference Center, San Francisco, California, iPRES 2009:the Sixth International Conference on Preservation of Digital Objects, California Digital Library, UC Office of the President, October 5-6, 2009, p. 184-185. Available at www.cdlib.org/services/uc3/iPres/presentations/SmithArmstrong.pdf (access 24 August 2011).

[43] Bak and Armstrong, p. 287. In 2008, LAC identified the core or minimum set of metadata required to transfer electronic records from RDIMS to the TDR. This metadata core set is influenced by the findings of

Post-Ingest Activities

Upon ingest, electronic records would undergo a series of manual and automated processes in the first phase of the LAC TDR: the Virtual Loading Dock (VLD). Here records would be scanned by antivirus software, parsed by metadata extraction tools such as JHOVE and DROID used to identify and validate file formats, subjected to checksum algorithms to establish integrity control, and consolidated within SIPs based on the file classification codes originally issued by RDIMS.[44] JHOVE extracted descriptive and technical metadata is then structured according to the Metadata Object Description Schema (MODS) and the Preservation Metadata: Implementation Strategies (PREMIS) standards upon which time a Metadata Encoding and Transmission Standard (METS) manifest is generated to collate all SIP content data to provide a comprehensive account of the processes electronic records have undergone throughout ingest. As conceptual carriers, the METS and MODS manifests would subsequently be used to transport metadata throughout successive stages of the LAC TDR. Technically validated SIPs held in the VLD digital storage area may be previewed by archivists at this stage where the augmentation of descriptive metadata may also take place.

As a SIP progresses through the TDR infrastructure, an unaltered preservation master copy of its digital objects and metadata can be stabilized as an Archival Information Package (AIP) which may then be assigned a unique and persistent identifier such as an Archival Resource Key (ARK) to ensure it may be easily located and retrieved

---

the InterPARES Project :I Authenticity Taskforce, ISO 15489:2001 (records management), and ISO 23081:2006 (managing metadata for records).

[44] JHOVE provides functions to perform format identification (what format the digital object is) and validation (what format the digital object purports to be). DROID performs the same functions but handles a much larger range of formats. Using both JHOVE and DROID ensures digital objects are properly identified and validated.

from TDR storage environments.[45] An AIP is then monitored for file format obsolescence and if necessary, migrated to different logical carriers according to a digital format registry such as the LAC LDFR.[46] All of these preservation activities may then be tracked and documented in a METS manifest kept within the AIP. Records metadata generated by the original creator, augmented by archivists, and accumulated during TDR processing and storage can then be mapped to the LAC descriptive metadata management and resource access system called MIKAN.[47] Once described in MIKAN, users may search processed electronic records through the LAC online interface and submit a request for individual records at which time a DIP is generated and exported from the TDR. The stable connection between the LAC TDR (preservation) and MIKAN (access) is made possible by the ARK identifiers assigned to each individual electronic record within an AIP.

Assessment of LAC TDR Approach

Upon ingest, there is theoretically little difference in the means by which government e-records and personal digital records are preserved and accessed within the LAC TDR. Put differently, a LAC TDR-compliant SIP derived from an individual donor would be subjected to extraction, technical validation, migration, and dissemination procedures comparable to those performed on a SIP derived from a departmental office within the Government of Canada. Moreover, private-side AIPs within TDR storage would command the same preservation monitoring as government-side AIPs while at the

[45] The LAC TDR presumably uses Redundant Array of Independent Disks (RAID) storage but this information could not be confirmed.
[46] See Library and Archives Canada, "Digital Initiatives at LAC, Digital Policies, Guidelines and Tools," available at http://www.collectionscanada.gc.ca/digital-initiatives/012018-2200-e.html (accessed 30 September 2011).
[47]LAC provides access to its published digital materials through the AMICUS public search tool.

same time benefiting from the commitment to long-term preservation and access given the proposed sustainability of the LAC TDR project. There is a glaring disparity however, in the pre-ingest environments of private and government records creators. For example, EDRMS, GC RMAP, and records managers, the front-end of the LAC TDR, cease to exist in the computing environments of individuals and so there is no conceivable way that personal digital records can meet even the minimal core metadata requirements deemed mandatory for TDR ingest. What is more, compliance with LAC endorsed metadata application profiles in addition to classification plans, file naming conventions, and RDAs cannot be expected of personal records creators unless these people are first identified and approached by archivists at the point of early records creation.[48]

A key strength of the LAC TDR approach lies in its pre-ingest metadata creation activity. For example, the RDIMS software application allows users to add metadata, such as name, description, type, and creator, to email, presentations, spreadsheets, and word processing files through a "document profile" interface.[49] RDIMS also allows users to document important contextual relationships between multiple digital records with its "Related" function. This type of pre-ingest activity minimizes the amount of retro-active labour archivists must perform in terms of arrangement and description as creator generated metadata pushes a significant portion of these responsibilities back to the records creators themselves. Paradoxically, the benefits yielded by LAC TDR pre-ingest metadata standards such as rich provenancial information and the documentation of

[48] The close equivalent of a RDA in the personal realm would be a donor or deposit agreement between LAC and an individual.
[49] Government of Canada, Office of the Information Commissioner, *Records, Document and Information Management System (RDIMS) User Guide*. Available at http://www.oic-ci.gc.ca/eng/rr-sl-odi-adi.aspx (accessed 25 December 2011).

relationships between records are somewhat lost when applied to personal archives as this level of requisite information may become too great of an obstacle, or simply too burdensome for private creators in pre-custodial environments and perhaps too labour intensive for archivists to retroactively construct before ingest.

A more pragmatic advantage of the LAC TDR approach, as with other large scale digital repository architectures, is their ability to secure sustainable funding from stakeholders with a vested interest in good recordkeeping and archival preservation of records with business value. Regular funding and other resources are paramount given that the fluidity of the digital landscape requires constant upgrading of technology and warrants the modification of policy and procedure as dictated by evolving cultures of corporate and government records creation. For example, the National Archives and Records Administration (NARA) has received nine-hundred-ninety-five million dollars since 2005 to fund its Electronic Records Archive (ERA) for United States Federal Government records.[50] Similarly, the LAC TDR project has received an estimated six million dollars in its first four years of development [51] In addition to this financial element, the policy and procedural infrastructure of the TDR allows, to some extent, archivists to form and enforce agreements and commitments made between records

[50] Alice Lipowicz, "White House attention nudges e-archive toward completion by September 2011," *Federal Computer Week* (August 26, 2010). Available at http://fcw.com/articles/2010/08/26/white-house-nudges-electronic-archives-to-completion-by-september-2011.aspx (accessed 24 August 2011).
[51]Library and Archives Canada, Preliminary Survey of a System Under Development Audit of the AMICAN Catalytic Initiative (June 2007). Available at http://nlc-bnc.ca/obj/012014/f2/012014-205-e.pdf (accessed 24 August 2011). The LAC TDR was part of a joint digital initiative with the LAC AMICAN system built to support management of and access to LAC's holdings. Smith and Armstrong, "Preserving the Digital Memory of the Government of Canada," p. 182.The LAC TDR was funded by the Treasury Board of Canada Secretariat (TBS) from 2007 to 2010. The TBS also supported the purchase of RDIMS licenses for all Government of Canada departments. See Robert Coffin, "Information Management and e-Government in Canada: From Government On Line to Service Transformation," Online presentation available at http://www.docstoc.com/docs/71273777/Developing-GOL-Funding-Options (accessed 24 August 2011); See also Government of Canada, "Libraries and Archives Canada, 2010-2011, Report on Plans and Priorities," pp. 1-34. Available at http://www.collectionscanada.gc.ca/012/012-205-e.html (accessed 24 August 2011).

creators and LAC.[52] The transfer of archivally-sound electronic records from government departments is considerably aided by intervention by records managers and archivists while policy and standardization documents, such as the Trustworthy Repositories Audit and Certification (TRAC) Checklist, help ensure proper long-term preservation and access. Library and Archives Canada may come to incorporate privately generated digital records into TDR workflows, but has yet to engage in the necessary pre-ingest research to determine what level of metadata granularity is appropriate for personal digital records and what degree of metadata creation can reasonably be expected of private individuals. The second OAIS-based methodology examined in this section provides a number of prospective approaches to influence the pre-ingest environments of personal records creation.

The PARADIGM Project

> Collecting archivists tend to distance themselves from the processes of records creation and management, which is often viewed as the remit of the records manager, and in the case of personal archives the remit of the records creator alone.[53]

A shortcoming of the LAC TDR approach to digital personal records is the absence of pre-ingest agents and activities comparable to those found in government record creating environments. Indeed, government records archivists have for some time been aware of the dividends paid by intervening in records creating environments. This interventionist approach to digital archives within the realm of the personal is the

---

[52] It should be noted that in a 2007 survey of forty-seven Government of Canada departments and agencies, it was found that while close to sixty-five-percent of responding areas used an EDRMS, the employees of these departments did not register electronic records in the system. It was estimated that only half of the records of archival value are captured in EDRMS installations. See Robert Nahuet, "The Management of Textual Digital Archives," pp. 37-38.

[53] Susan Thomas and Janette Martin, "Using the Papers of Contemporary British Politicians as a Testbed for the Preservation of Digital Personal Archives," *Journal of the Society of Archivists* 27:1 (April 2006), p. 36.

backbone of the Paradigm project, a joint initiative of the Bodleian Library, University of Oxford and the John Rylands University Library, University of Manchester to evaluate the issues involved in the curation of born-digital personal archives from ingest to access. The project's research and key output adheres to the OAIS model in that it is mapped to its functional entities while at the same time adheres to the more traditional archival workflows of acquisition, appraisal, preservation, and access.[54] The project chose the digital personal papers of six contemporary British politicians as its test-bed largely because it "targeted an area of importance to both institutions and raised the profile of digital preservation issues with participating politicians."[55] As it is beyond the scope of this thesis to synthesize all of the research findings disseminated by the Paradigm project, the following section examines the pre and post-ingest activities of Paradigm with a special emphasis on its instructions for early archival intervention.

Pre-Ingest Activities

There are a number of preliminary steps archivists may take before they attempt to acquire personal records and submit them for ingest to an OAIS digital repository. Chief among these is the fostering of an ongoing relationship between the archivist and the records donor at the early stages of records creation. Following the establishment of a sustainable archivist-donor relationship, Paradigm suggests the provision of regularly

---

[54] Source material for the analysis of the Paradigm project is derived from the *Paradigm: Workbook on Personal Digital Archives*, and supplemented by Susan Thomas, *A Practical Approach to the Preservation of Personal Digital Archives: Final Report to the JISC*, (Oxford: Bodleian Library, 2007), pp. 1-38; and the Paradigm project website available at http://www.paradigm.ac.uk/index.html (accessed 15 July 2011).

[55] Thomas, *A Practical Approach to the Preservation of Personal Digital Archive* p. 10 and 16. The six participating politicians were drawn from the Conservative, Labour, and Liberal Democrat parties and included Members of Parliament, Peers and Members of the European Parliament. The project scope was slightly extended to include a website harvesting project and to work with older digital materials of former Cabinet Minister Barbara Castle held at the Bodleian Library. While it may be argued that the archives of politicians are hardly representative of personal digital archives throughout society or that the records within them serve only to document the professional persona of individuals, Paradigm is one of the only, if not the earliest, comprehensive example of archivy engaging the pre-custodial environment in earnest to determine practical solutions to well-known problems with digital personal archives.

updated practical guidelines and tips to individuals for ongoing maintenance of their personal archives. According to these best practices, records creators are encouraged to: organize and name files appropriately, select suitable/non-proprietary software when possible, backup and synchronize data across multiple on and offline platforms, and make records self-documenting by adding metadata.[56] Here, producer-generated metadata such as title, location, size, dates last accessed and modified is often automatically created by software and hardware applications, while metadata such as author(s), purpose, document version, keywords, or free-text description must be manually entered in the "properties/summary" option for individual files or kept in a simple text file within a folder of multiple files.[57]

A second important preliminary step involves archivists surveying personal digital collections to establish their content and context (metadata), which is also used for the purpose of managing these collections as SIPs and AIPs. This survey may take the form of a questionnaire to elicit information on the on and offline location(s) of original and duplicate digital records, the types of records in both paper and digital forms, the frequency of recordkeeping and preservation activities, in addition to privacy and intellectual property concerns.[58] In addition, Paradigm suggests archivists perform screenshots of personal computers to capture graphical representations of system information (such as desktop organization, software application icons, and email clients) and capture more detailed file directory structures of personal computers that may be

[56] *Paradigm: Workbook*, pp. 281-282. Cited formats include: Open Document Format (ODF) and PDF/A for textual records, MySQL for databases, TIFF format for raster images, and MBOX for email.
[57] Ibid., pp. 277-288. Paradigm also notes records creators should comply with World Wide Web Consortium (W3C) recommendations and validate their personal websites, backup files, store copies off-site, and administer their system hardware and software on a cyclical basis. Other tips provided here include: using passwords and open-source encryption software, intellectual property and copyright awareness, and advice on handling digital legacy records.
[58] Ibid., pp. 24-26 and pp. 289-291.

accomplished via its command line interface.[59] The Paradigm team concludes that deposit and gift agreements should be signed by donors or depositors up front in order to establish the ownership of the digital collection, to identify third party copyright holders, to receive permission to undertake preservation actions (migration and backup), and to establish access permissions as well as any restrictions. The pre-acquisition appraisal is intended to assess the content, context, structure and technical state of the creator's archive before the archivist attempts to gain intellectual control of its records.

The third major pre-ingest component of Paradigm involves five potential acquisition strategies deployed to capture digital records for eventual OAIS intake. In the first proposed strategy, archivists perform on-site "snapshot" captures of digital records held by creators which are then immediately accessioned in a digital accessions repository. These accessions adhere to the agreement made between the institution and the creators to determine the types of records that can be acquired during snapshots and to ensure they occur at regular predetermined intervals (*Regular snapshot accessions*).[60] In a second strategy Paradigm proposes that archival institutions engage in a continuing relationship with records creators and provide management oversight in the form of advice and best practice guidelines. In this strategy, digital records are not acquired through regular accessioning but remain in creator custody until such time as they may be appraised and transferred to an archival repository (*Post-custodial approach*).[61] In a

---

[59] Ibid., pp. 26-30. See also the *Paradigm: Workbook, Capturing directory structures* online available at http://www.paradigm.ac.uk/workbook/record-creators/capturing-directory-structures.html (accessed 16 June 2011). Other pre-ingest activities include the creation of transfer lists and the migration of records to removable media such as CD or USB drives. Snapshots of personal Websites and blogs are also recommended using HTTrack software which captures the site's file content and stores it on a local hard drive.

[60] *Paradigm: Workbook*, p. 11.

[61] Ibid., pp. 12-13. A combination of the snapshot and post-custodial approaches is also proposed as this would allow archivists to make regular accessions of pre-appraised records thereby facilitating their

similar strategy, archivists maintain a relationship with creators as in the two previously mentioned approaches, but only acquire records once the hardware or software on which they reside is no longer used by creators. Here, the archivist acquires retired personal computers, hard disk drives, physical storage medium, peripheral devices, as well as software applications during the lifetime of creators (*Transfer via retired media*).[62] A fourth strategy also hinging on the archivist-donor relationship involves the provision of digital services to individuals such as email and live chat, remote storage and data backup, in addition to web hosting and personal digital assistants. Here, Paradigm advances a strategy where individuals create records as they please but do so according to the tools and services designated by the archival institution and would eventually see records creators uploading their records to an OAIS–based repository remotely (*Self-archiving*).[63]

Acknowledging that institutions will invariably continue to acquire digital personal records much in the same way they do paper-based genres, Paradigm suggests that the traditional method of capturing materials from creators at the near-end of their careers/lives will likely exist alongside the other four approaches stated above (*Traditional approach*). In such cases, Paradigm offers possible solutions to ameliorate the severity of file format obsolescence and inaccessible storage media. For example, hard disk drives may be removed from personal computers to have their internal files and software extracted and copied via disk imaging techniques while corrupted or unreadable digital information may often be captured using proprietary and open-source data

---

ingestion and management within a controlled OAIS environment while other records not yet ready for acquisition remain in creator custody guided by the archivist(s) responsible for their eventual transfer (*Combining snapshot and post-custodial approaches*).

[62] Ibid., p. 14.

[63] Ibid., p. 15.

recovery tools or by similar workaround procedures offered by third party services (*Digital forensics and archaeology*).[64]

Post-ingest Activities

Much like the LAC TDR approach, Paradigm employs a number of digital services within isolated on and offline environments following OAIS ingestion of digital records. In the first phase of records intake, a standalone (offline/non-networked) accessions environment is used for the initial processing of SIPs to forestall malware attacks and other potential security breaches. Here, backup and working copies are generated and virus checks are performed, while digital objects are further subjected to metadata extractors as well as file format identifiers and validators.[65] Persistent identifiers (PIDs) are then assigned to individual digital items according to a specific scheme at which point a METS manifest containing extracted and manually entered metadata is generated to accompany those digital items through the construction of an AIP.[66]

In the second phase of OAIS intake, AIPs are stored within a separate digital preservation repository architecture which may be referred to as a "dark archive" in that it is "designed purely for archival storage that should be subject to preservation monitoring and actions, and secured to protect material contained in it that may be sensitive and

---

[64] Data extraction removes files from fragile media at which time archivists migrate the data to a more stable format for processing. This process obviously becomes more difficult when data extraction involves recreating the entire antiquated computing platform (OS, disk drive peripherals, and obsolete software applications).

[65] Thomas, *A Practical Approach to the Preservation of Personal Digital Archives*, pp. 26-28. Tools and registries used by Paradigm include: DROID (format identification), PRONOM and FILEXT (format registries), JHOVE (format validation and technical metadata generation), and the National Library of New Zealand Extraction Tool (metadata extraction). If an identified file format is not supported by the repository, it is subsequently converted into a format that is.

[66] Ibid., pp. 25-32; and *Paradigm: Workbook*, pp. 48-140. Persistent Identifier schemes include: Uniform Resource Name (URN), Persistent Uniform Resource Locator (PURL), Handle System, Digital Object Identifier System (DOI), and Archival Resource Key (ARK). PIDs must be resolvable, meaning they must provide information on how to access the object they identify.

subject to embargo."[67] AIPs contain important technical and preservation metadata adhering to the PREMIS Data Dictionary which defines the core semantic units that digital repositories should know in order to adequately perform their preservation functions.[68] As the ongoing authenticity and integrity of digital records is important to both archivists as well as original creators, Paradigm applies measures to create fixity information through cryptographic hash functions (MD5 or SHA-1), cyclical redundancy checks (CRC 16 or 32), and/or digital signatures that encrypt hash values through private and public keys.[69] All metadata generated by processes within the dark archive is consistently recorded in METS manifests that are inextricably linked to their associated digital objects. Once access restrictions and content liability expire, access copies (DIPs) may be generated and published to a separate online repository while master copies remain in the preservation environment.

Assessment of the Paradigm Approach

> Early intervention is an important principle for digital archivists, but it is relatively new to see this kind of relationship between archivists and creators of personal records. It turns the archivist's relationship with a depositor on its head.[70]

The significance of the Paradigm project is conveyed in its interventionist approaches to personal digital archiving practice. While this type of approach was first proposed theoretically by Adrian Cunningham in the 1990s, this technique "has not been so well explored in practical terms, or at least this exploration has not been well

---

[67] Thomas, *A Practical Approach to the Preservation of Personal Digital Archives*, p. 33. Paradigm chose the Fedora repository system software for its prototype digital preservation repository over the DSpace software system.
[68] Priscilla Caplan, *Understanding PREMIS*, (U.S.A.: Library of Congress, 2009), pp. 4-5. Available at http://www.loc.gov/standards/premis/ (accessed 25 August 2011).PREMIS does not specify how these core semantic units are to be recorded in, for example, an XML-based metadata schema.
[69] *Paradigm: Workbook*, pp. 152-156.
[70] Thomas and Martin, "Using the Papers of Contemporary British Politicians," p. 38.

documented for the benefit of the larger archival community."[71] However, much like Cunningham's pre-custodial intervention approach, the Paradigm project is subject to criticism. By intervening in the records creating practices of an individual, Paradigm unavoidably assumes and confers a degree of historical or cultural significance upon that person which may or may not be warranted. Moreover, through such intervention Paradigm's approach undermines key tenets of foundational archival theory and may be viewed as a disruption of natural or organic processes of creating archives. Yet, as Paradigm researchers, Susan Thomas and Janette Martin state "[d]espite these philosophical issues, the project team decided that the vulnerability of digital records, to accidental or deliberate loss, merited a compromise of principles."[72] Philosophical criticisms aside, there remain a number of pragmatic reasons for archivists to support the five acquisition strategies proposed by the Paradigm project while at the same time being aware of possible obstacles to their implementation.

Frequent and well-planned snapshot accessions of personal digital records require regular visits by an archivist either onsite or remotely. These snapshot accessions therefore allow archivists to perform value appraisals of records closer to the time of their actual use, and provide the archivist with an opportunity to consult the creator regarding additional contextual information which may inform the creation of technical and preservation metadata in addition to future archival description processes. The drawbacks of this approach, however, become evident with the potential for considerable duplication of digital records which may occur through repeated accessions. In addition, if accessions

[71] Thomas, *A Practical Approach to the Preservation of Personal Digital Archives*, p. 9. Thomas does acknowledge the work of Cunningham as well as Lucie Paquet as notable exceptions in the area of personal archives.
[72] Ibid., p. 38.

are not properly timed, they may miss important records created and deleted between snapshots. Post-custodial approaches to personal digital archives delay preservation within the digital repository while at the same time allow archivists to provide advice on records creation and avenues of preservation. Yet unless there is some form of legal agreement established up front, the private records creator is under no obligation to transfer digital records from their custody to a repository, thus rendering the time and resources spent on assisted digital curation fruitless.

Pro-active acquisition of digital records via retired media sees digital records arriving at the repository with their supporting hardware and software elements intact, which is often required to render the record's data content in addition to its contextual and structural attributes. This approach may also make any required data extraction or data recovery from fragile media less arduous if the archivist can survey the platforms used throughout the lifetime of a creator. Providing online tools and services to records creators and supplying donors with channels for remote uploading to an accessions repository is a captivating prospect for archival institutions seeking to bring personal digital records under their control as early as possible. This self-archiving approach benefits creators substantially, given that they have a safe and secure repository in which to store their records in the event of a system crash, malware attack, and through inevitable upgrades to platforms within their personal computing environments. This approach does run the risk of becoming more of an active records management repository where digital records are deposited one day only to be retrieved the next. What is more, the ongoing provision of archivally-sanctioned file formats, tools, and services may also

be too great of an imposition given the propensity of individuals to choose software and services that facilitate their unique means of records creation.

In closing, the principal investigators of the Paradigm project readily note the advantages and disadvantages of the acquisition methods mentioned above. They argue that a flexible combination, as opposed to a rigid implementation, of snapshot, post-custodial, self-archiving approaches along with transfers of retired media form in all likelihood the most pragmatic acquisition strategy for personal records in the digital era.[73] Pre-custodial intervention, as articulated by the Paradigm project is undoubtedly still in its infancy stages and it does require an extensive reorientation of how archivists think about digital acquisition and preservation while at the same time placing new demands on archival institutions in terms of hardware and software applications required for the actual processing of personal digital records in OAIS repositories. The Paradigm and LAC TDR examples place great emphasis on metadata generation, virus scanning, fixity checks, file format conversion, preservation monitoring, repository architecture and many other technological components of OAIS environments. Yet, these two models do not discuss the possible scalability of their respective approaches to archival institutions that acquire the digital records of private persons but may lack the financial and technical resources to make this level of digital archiving a reality.

Archivematica

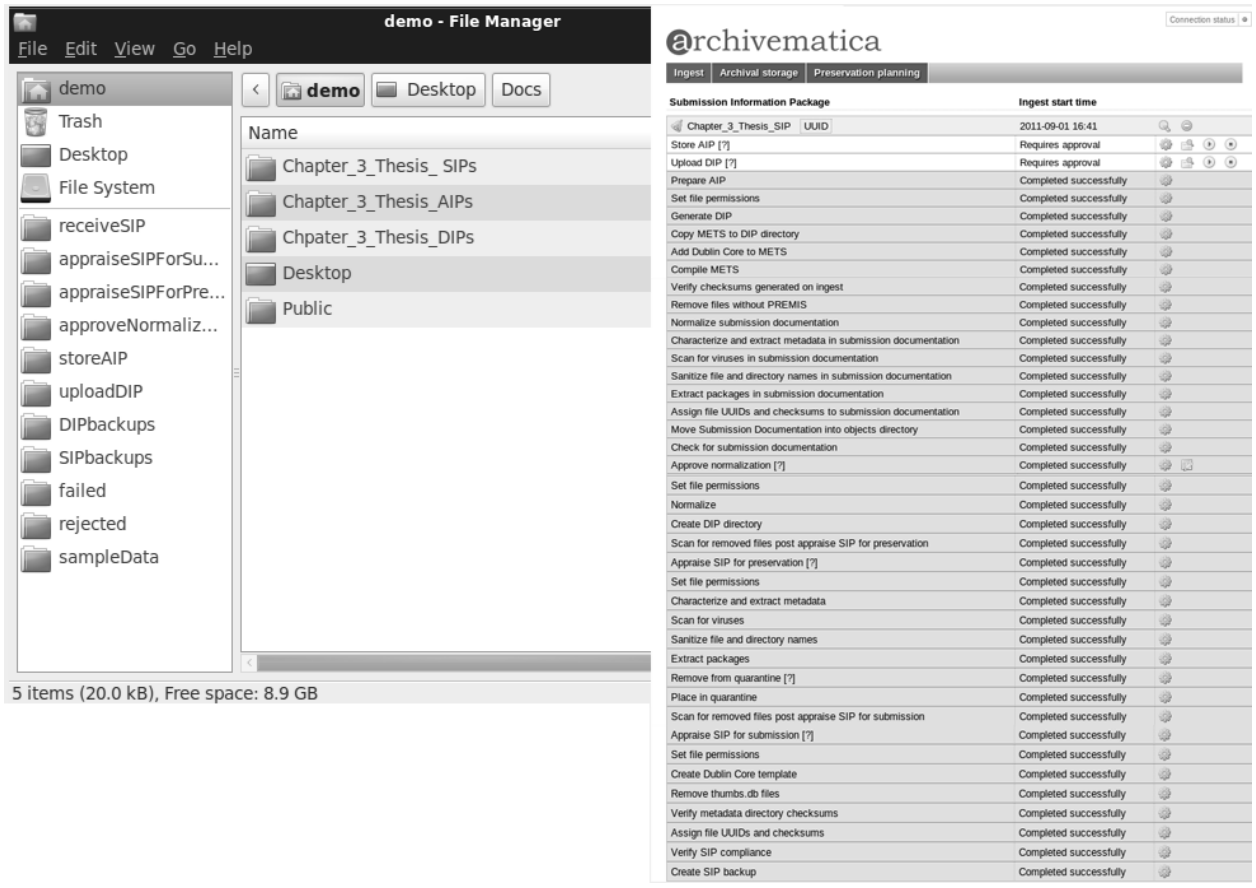
Developed by Canadian technical services provider, Artefactual Systems Inc., Archivematica is an OAIS-compliant comprehensive digital preservation system using a micro-services design pattern to provide a suite of open-source tools that allows

[73]Thomas, et al. *Paradigm: Workbook*, pp. 10-16.

archivists to process digital objects from ingest to access.[74] Archivematica may be installed on any personal computer operating system via virtualization software such as VirtualBox or VMware and is synchronized with the file system of the host personal computer to create a manageable preservation environment capable of processing a number of contemporary file formats. Provided that pre-custodial approaches such as those advocated by the Paradigm project are successful in capturing personal digital records in the present, Archivematica may be used as a tool to process those materials for ongoing archival preservation.

[74] Source material for this section is derived from the Archivematica project wiki site, available at http://archivematica.org/wiki/index.php?title=Main_Page (accessed 30 August 2011); video and PowerPoint presentations created and published online by Artefactual Systems Inc. team members; and testing of Archivematica 0.6 and 0.7 alpha releases installed and used on a Windows XP OS with VirtualBox software. This testing occurred between September 2010 and August 2011.

Figure 3.3 Archivematica 0.7.1 – alpha [75]



Pre-ingest Activities

Actual pre-ingest activities of Archivematica are limited due to its function as a technological device which invariably relies on human agents to first acquire then submit digital objects to the system environment before processing may be initiated. The early stages of Archivematica workflows involve the archivist accepting digital objects and metadata from records creators, which is followed by the transfer of those materials to

[75] Figure 3.3 combines two screen shots of the Archivematica 0.7.1-alpha file manager and dashboard interfaces. Figure 3.3 adheres to the terms of the Creative Commons Attribution-Share Alike 3.0 Unported (CC BY –SA 3.0) license. Available at http://creativecommons.org/licenses/by-sa/3.0/ (accessed 25 August 2011).

Archivematica through a shared host-virtual appliance directory or directly through ingest channels such as a SFTP client.[76] At present, Archivematica is capable of ingesting a number of media including: audio, email, open office XML, plain text, portable document format, presentation files, raster and vector images, raw camera files, spreadsheets, video, and word processing files.[77] Although the Archivematica environment does not require that digital records meet any standard of metadata for transfer or intake, any technical, descriptive, and preservation metadata generated by creators may accompany a set of digital records upon ingest and be subsequently augmented by an archivist.

Post-ingest Activities

Upon ingest, SIPs are processed through six main areas and subjected to a series of self-contained micro-services. In the first phase of processing, a folder of digital objects is formatted for SIP compliance, a checksum is issued for all objects, while logs and metadata sub-directories are created to house processing information and metadata accumulation.[78] A structured SIP is then manually moved to the *receive*SIP stage in the file manager interface at which point it is then drawn into the Archivematica pipeline and appears in the web-based dashboard that the archivist uses to monitor and control the

[76] Here the PC acts as a centralized workstation for ingest where digital objects are copied from physical carriers such as flash drives, CDs, or DVDs or received via email. Once received, the archivist imports digital objects to Archivematica adhering to the client-server relationship where materials are transferred from a PC (client) to the virtual environment (host-server). FileZilla and WinSCP FTP Client software was used to test this procedure. Previous versions of Archivematica allowed users to link shared folders between both the host and the virtual environments by amending the python programming language scripts in the virtual machine's terminal interface. Hypothetically, individual records creators could bypass the archivist and import their digital records directly to an Archivematica installation using the same SFTP ingest channel.

[77] Archivematica supports preservation planning for over fifty different file format extensions.

[78] The metadata subdirectory contains a blank Dublin Core XML template which the archivist may populate manually as well as a submission documentation folder for donor agreements or accessions information.

suite of micro services (see figure 3.3) .[79] Next, the archivist is prompted to appraise the SIP contents and approve it for submission to the quarantine stage where it is scanned for viruses for a pre-determined period of time.[80] Passing the virus health check the archivist is again prompted by the program, this time to release the SIP to further micro-services such as extraction of zipped content, file and directory name sanitation (removal of prohibited characters), file format identification and validation against external specifications using the File Information Tool Set (FITS) also used to extract technical metadata, which is then added to a PREMIS manifest within the SIP.[81] At a third checkpoint, Archivematica requires the archivist to approve the SIP for preservation. At this stage, digital objects within the SIP are "normalized" – meaning preservation and access copies are automatically generated and maintained along with the original file formats. Here, the digital objects, their associated metadata, and a processing log are repackaged as an AIP and collated with a METS wrapper to facilitate transfer.[82] Next, the archivist approves the uploading of a DIP to an online access repository and then approves the transfer of the AIP to interface with a variety of possible digital storage locations.[83]

---

[79] Archivematica generates a SIP backup folder and assigns PIDs in this first stage. Digital objects are moved throughout the forty micro-services according to the UNIX pipeline design pattern where standard output of one micro-service initiates the input of the next micro-service.

[80] Archivematica is typically configured to be online. This means that its CLAM AV tool is frequently updated with virus definitions. If SIPs do not pass the virus check, they are removed to an isolated directory.

[81] FITS was created by the Harvard University Library, Office for Information Systems. Tools within this set include: JHOVE, ExifTool, National Library of New Zealand Metadata Extractor, DROID, and FFIdent.

[82]The Library of Congress BagIT format is used to zip AIP content.

[83]The AIP may be stored on the HDD of the personal computer, on a network drive, on an external HDD, or in a Fedora digital repository. The DIP is most easily uploaded to an instance of the International Council of Archives Access to Memory (ICA AtoM) descriptive database. Directing these AIP and DIP transfers is accomplished by amending the python programming language scripts using the Archivematica terminal. These two processes were successfully tested by the author in the Archivematica 0.6-alpha release.

Assessment of Archivematica

As discussed in chapter one of this thesis, responsibility for the acquisition and preservation of personal digital records in Canada is distributed across many archival repositories at the national, provincial, university, and municipal levels. In many cases, however, archivists in mid-size institutions acquiring personal digital records lack the technical and financial infrastructure to engage in a comprehensive digital preservation project and may view the OAIS reference model as methodology reserved solely for organizations such as Library and Archives Canada or the National Archives and Records Administration. In view of these circumstances, OAIS-based Archivematica installations represent a guided and pragmatic approach to archiving personal digital records in the present. What is more, as a highly scalable application, Archivematica may be run as a bootable USB key in small computing environments or customized in a virtual or bare-metal installation to interface with pre-existing digital preservation architectures in university settings. For the purposes of archiving personal digital records, however, the true strength of the Archivematica project lies in its media-type preservation plans and its corresponding normalization tools.

The Archivematica project has chosen normalization as its primary preservation strategy whereby media-types (audio, video, email, etc.) are transformed into specified preservation formats based on contemporary best practices, and specified access formats based on the ubiquity of viewers (software applications) for that file format. Moreover, both preservation and access normalizations adhere to an open-source standard and rely on non-proprietary tools to perform these format transformations upon OAIS ingest.[84]

[84]For example, audio media-type file formats such as AC3, AIFF and WMV are converted to the WAV preservation format and the MP3 access format. The tool used to perform these normalizations is the open-

While it is possible for archivists to perform these format transformations manually through individual normalization tools, Archivematica automates this process and the archivist need only intervene upon a normalization failure or in the absence of a default normalization tool.[85] Built-in media-type preservation plans guided by best-practice groups such as PRONOM also inform the archivist of changes in the risk status or migration channels through online updates to the Archivematica system. Furthermore, as the original format of a digital object and its preservation and access surrogates are maintained in tandem, normalization inherently supports other migration and emulation preservation strategies which require access to the pre-conversion format. In essence, Archivematica facilitates the construction of personal digital archives in the present by lowering the barrier to effective processing of contemporary file formats through its highly automated and robust micro-services architecture.

While a combined methodology based on the best elements of the LAC TDR approach, the Paradigm project, and the Archivematica software may assist archivists in the acquisition and preservation of personal digital records relative to their generation, there will inevitably be those records creators that have been bypassed or otherwise overlooked by coeval documentation strategies and whose records will presumably require some degree of digital archeology. Accordingly, front-end approaches to personal digital archives will nevertheless be coupled with parallel archival practices such as those epitomized by the Emory University and the Harry Ransom Centre. Yet, archivists must still look to the future and strategize for future acquisitions of personal digital records as

---

source FFMPEG multimedia conversion tool. In short, normalization involves the original format, the preservation format, and the access format of a digital object.

[85] It should be noted here that as Archivematica is still undergoing important stages of development, there are a number of failures that may occur between ingest and access. That said, it is easy for a user to see what may have caused a system failure by examining error logs.

well as the ongoing preservation of those digital materials already within their care. The concluding section of this chapter examines exceptionally progressive approaches to personal digital archives as well as the new and emerging technological tools and services used within.

Embryonic Strategies for the Future of Personal Digital Archives

Among the many innovative approaches offered by the Digital Lives research project, iCuration stands out as a particularly advantageous front-end utility for both personal records creators and archivists. Typically, archival work in the areas of appraisal, donor consultation, and records acquisition is performed on site or put differently, offline. Digital Lives proposes that much of this archival work can be taken online in what the project refers to as iCuration where archivists may provide online advice and training to donors, issue suitable tools and services for individual management and preservation of digital materials, and capture personal digital records remotely.[86] In point of fact, a great deal of iCuration may actually be accomplished today through a basic remote desktop session which would allow an archivist to survey a donor's personal computer(s), examine and document recordkeeping behaviours, as well as appraise and even acquire personal digital records via SFTP or SSH transfer or through a user-friendly file hosting service built along lines similar to proprietary cloud-based services such as Dropbox.[87] Moreover, online training and advice can, at present, be provided by way of

[86]Jeremy Leighton John et al. "Digital Lives, Personal Digital Archives for the 21st Century: An Initial Synthesis, Beta Version 0.2," (March 2010). Available at http://britishlibrary.typepad.co.uk/files/digital-lives-synthesis02-1.pdf (accessed 21 April 2010), pp. 124-125.

[87] Remote Desktop Protocol (RDP), designed by Microsoft Corporation, provides remote display and input capabilities over network connections. Remote desktop clients (available for both Windows and Mac operating systems) allows users to access an entire personal computer from any other computer (and more recently smartphones) supporting RDP. Using RDP, an archivist could interact with a donor's computer in the same way IT support interacts with an employee workstation.

RSS feeds which may also communicate changes to "terms of use/service" for email, social media platforms, and other online service providers the creator has employed for file storage.[88]

By engaging with individuals in this kind of networked relationship, archivists may tailor iCuration for a diverse demographic of records creators with varying levels of computing expertise. As iCuration aims to assist individuals to create, maintain, and sustain their digital archives, this concept may be extended to include the provision of specific technologies to personal records creating environments. For example, the Home and Office Painless Persistent Long-Term Archiving (HOPPLA) software application developed at the Department of Software Technology and Interactive Systems, Vienna University of Technology may prove to be a valuable asset for personal digital archives maintained in creator custody.[89] Built on client-server architecture and following a service model similar to that of firewall and antivirus software packages, the Hoppla research-prototype "combines back-up and fully automated migration services for data collections in small office environments" where client-side users provide information about their archives to an update-service which in return analyzes the content and issues information on how and under which circumstances digital objects should be migrated to more stable formats.[90] Tools such as ImageMagic, ps2pdf, and MEncoder installed on the client side host computer are then automatically called and executed by Hoppla where

---

See Microsoft Corporation MSDN Library, "Remote Desktop Protocol," available at http://msdn.microsoft.com/en-us/library/windows/desktop/aa383015(v=vs.85).aspx (accessed 4 October 2011).

[88] As webmail services and social media sites are constantly updating and changing their terms of use policies, RSS feeds can be used to warn personal records creators if their online content is in danger of being deleted or compromised. Digital Lives also provides an informal survey of the privacy and copyright statements of many contemporary online service providers. See John et al. "Digital Lives," pp. 88-96.

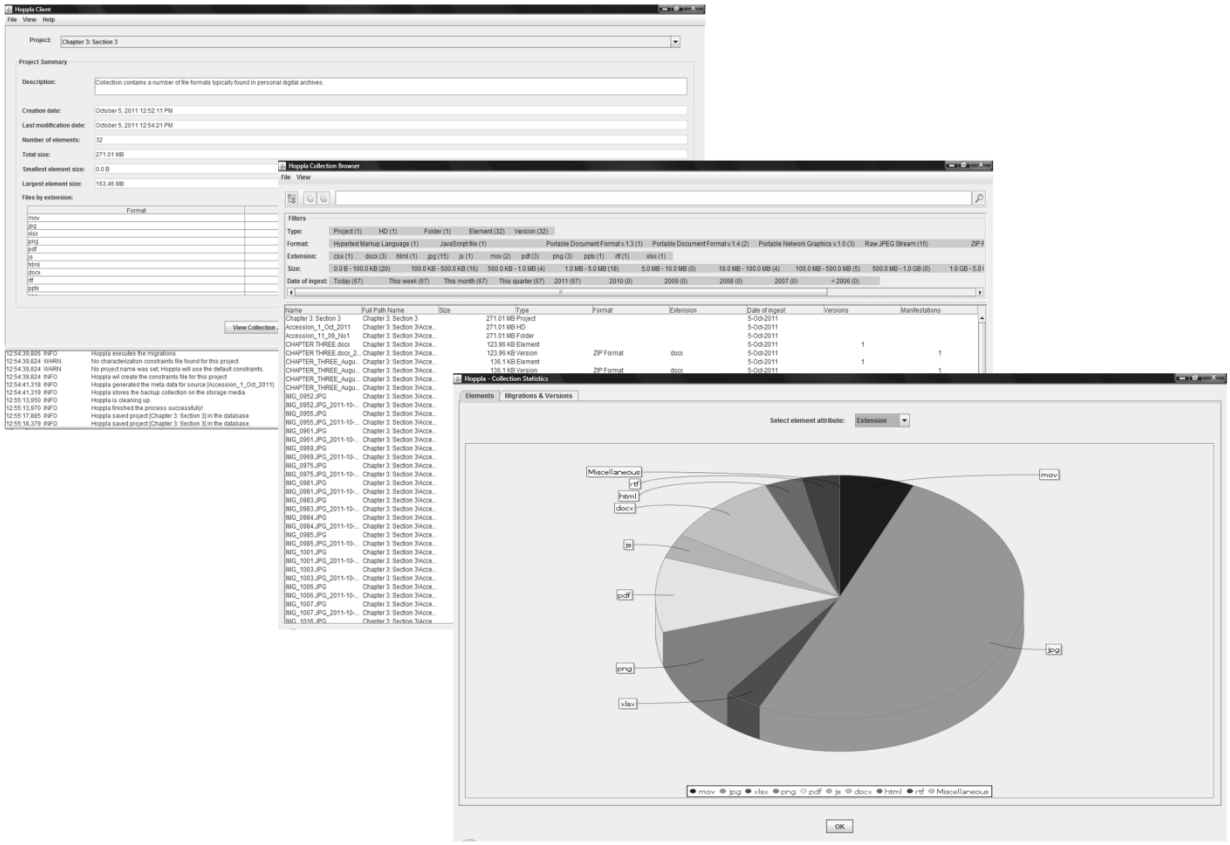
[89] Stephan Strodl et al., "Automating logical preservation for small institutions with Hoppla," available at http://www.ifs.tuwien.ac.at/dp/hoppla/ (accessed 25 September 2011).

[90] Ibid.

migrated digital objects are then transferred to internal or external storage media determined by the user. Theoretically, Hoppla could be configured to automatically request information from an update service at predetermined intervals set by the archivist and allow migrations to occur whenever digital objects were at risk.

In addition to the client-server and migration functionalities, Hoppla provides a collection browser with a number of different data visualizations (figure 3.4) to provide filtered and statistical overviews of the varying types and quantities of file formats within a personal digital archive. Though still in nascent stages of development, the current public release of Hoppla version-2.1 may be installed, configured, and used by personal records creators with relative ease and as such, provides an accessible and practical pre-custodial utility for personal records creators and personal records archivists alike.

Figure 3.4 Home and Office Painless Persistent Long-Term Archiving (HOPPLA) 2.1 User Interface[91]



> All archives use some form of metadata for description, reuse, administration, and preservation of the archived object. There are issues related to how the metadata is created, the metadata standards and content rules that are used, the level at which metadata is applied and where the metadata is stored.[92]

Metadata is integral to the affirmation of record authenticity, the discovery of record provenance, and the establishment of relationships between digital records. Yet, there is a

[91] Permission for Hoppla screen shots obtained from Petar Petrov, Information and Software Engineering Group, Institute of Software Technology and Interactive Systems, Vienna University of Technology. Additional technical information about Hoppla available at http://www.ifs.tuwien.ac.at/dp/hoppla/ (accessed 25 September 2011).

[92] Gail M. Hodge, "Best Practices for Digital Archiving: An Information Life Cycle Approach," *D-Lib Magazine* 6:1 (January 2000), p. 8.

considerable disconnect between required archival metadata and the people who must create it. For example, of the fifteen properties of the Dublin Core Element Set, Version 1.1, the majority of these fields in a schematic must be manually populated depending on the software application an individual is using to create a digital record. This point is illustrated in the creation of a typical word processing document where twelve Dublin Core elements require manual input by a user with the remaining three elements being created by the software application itself.[93] This is an important reality to be acknowledged as some archival metadata application profiles, such as GC RMAP or conceptual metadata schema informed by the InterPARES *Creator Guidelines*, consist of far more than the fifteen required Dublin Core elements, many of which are also not supported by automated processes.

Therefore, the responsibility of metadata creation invariably falls to the personal records creator or the archivist processing the digital collection which in both cases is tedious and extremely time consuming. What is more, human created metadata can be inconsistent, inefficient, and costly. In a study of the time and effort involved in self-archiving, researchers discovered that "The median time for metadata entry is 5 minutes and 37 seconds per paper" while a separate study examining the importance of quality assurance for metadata data estimated the re-editing of two-thousand-five-hundred metadata records by cataloguers cost "about 13 minutes or £2.60 per record."[94] In short,

---

[93]Jane Greenberg et al., *Final Report for the AMeGA (Automatic Metadata Generation Applications) Project*, (Chapel Hill: University of North Carolina School of Information and Library Science, 2005), pp.16-17. The three elements supported by automatic metadata creation are title, creator, and date. Using DROID, the required Dublin Core element "format" would be fulfilled.

[94] Leslie Carr and Stevan Harnard, "Keystroke Economy: A Study of the Time and Effort Involved in Self-Archiving," (University of Southampton, 2005) available at http://eprints.ecs.soton.ac.uk/10688/ (accessed 25 September 2011) and Jane Barton et al., "Building Quality Assurance into Metadata Creation: an Analysis based on the Learning Objects and e-Prints Communities of Practice," in Sutton, S. and Greenberg, J. and Tennis, J., eds. *Proceedings 2003 Dublin Core Conference: Supporting Communities of*

the granularity of metadata desired by archival institutions for digital records outstrips the ability of individuals and the software they use to create records.

In seeking to rectify this dilemma, there are two possible approaches. First, utilities may be developed to automatically generate metadata for a core set of software applications such as word processing documents, image files, spreadsheets, or audio and video files. These utilities may take the form of a customized desktop application which identifies and makes connections between personal records in, for example, the way Google Mail currently uses ranking algorithms based on keywords and send/receive frequency to classify certain messages as important.[95] By incorporating this type of algorithm design within a personal computer, relationships could conceivably be made between records containing similar keywords, dates, and file formats. This type of utility could also be informed by "fuzzy hash" techniques which search and identify homologous files having similar strings of binary data (hash values) but are not exact duplicates.[96] Archivists may also develop user-friendly metadata application profiles for personal digital records informed on archetypal recordkeeping and archiving behavioral patterns identified by PIM research. Here, an individual would be prompted to augment metadata within an attached manifest only for those records or groups of records which conform to their personal folksonomies of value (identity, memory, sentimental, or functional value).

Discourse and Practice - Metadata Research and Applications (Seattle: Information Institute of Syracuse, 2003) available at http://eprints.rclis.org/handle/10760/5237 (accessed 25 September 2011).

[95]See Google, Gmail, "How importance ranking works" available at https://mail.google.com/support/bin/answer.py?answer=186543 (accessed 25 September 2011).

[96]Fuzzy hashing can match inputs that "have identical bytes in the same order, although bytes in between these sequences may be different in both content and length." See Ssdeep website available at http://ssdeep.sourceforge.net/ (accessed 25 September 2011).

A second approach to automated metadata creation involves the extraction of information using complex software programs. As many genres of digital records, notably text-based files, conform to syntactical rules or certain conventions of representation, algorithms may be developed to search for specific information according to where it is typically found within that genre of record. Exploiting the structure of digital records through the parsing of specific areas may yield important technical and descriptive metadata that was not manually entered by the creator. Semantic metadata, or data within text-based files that is unstructured and not machine readable, requires special tools such as text-mining and optical character recognition (OCR) applications to identify and extract information. Again, by leveraging PIM research findings, archivists may come to have a better understanding of how individuals use certain keyword vocabularies to describe sets of record genres. In knowing what keywords to look for, archivists may then use text-mining to search for and extract semantic metadata which may prove useful in the discovery of relationships between digital records. Though more research and development of automated metadata generation is required before these two prospective examples may be realized in everyday practice, these ideas should figure prominently in the minds of archivists given the importance they place on metadata.

Summary of Chapter: No Longer a Poor Cousin in the Family of Archival Theory and Practice[97]

The once fallow field of personal electronic records scholarship has become a burgeoning area of professional research and development. For example, digital

[97] In his discussion on the history of personal archives theory and practice, Canadian archivist Robert Fisher notes "Private archives are a poor cousin to government archives in the family of archival theory. Archivists who work with private archives often lament the absence of a professional literature on which to base their work and thought about the archival endeavour." See Robert Fisher, "In Search of a Theory of Private Archives: The Foundational Writings of Jenkinson and Schellenberg Revisited," *Archivaria* 67 (Spring 2009), p. 2.

archeology as performed on the Rushdie and Joyce archives has set new benchmarks for the processes of personal data recovery and has set the stage for future explorations into digital forensic analysis of privately generated digital materials. The OAIS reference model, once reserved for the preservation of space data, has found its way into the Trusted Digital Repository of Canada's national library and archives and into the installation of Archivematica on personal computers of small archival institutions dispersed across the country. The Paradigm project's proposed acquisition methodologies and Digital Lives iCuration are the evolution of the ideas first put forth by Adrian Cunningham in the 1990s and are in many respects, the realization of pre-custodial intervention. Adding to this chorus of innovation in personal archives is the Hoppla system which, along with other systems that are sure to follow, promises to have a considerable impact on the way archives are maintained by private individuals in the future. While many of the technologies and procedures considered in this chapter are new -- with some being slightly idealized in that they have yet to be extensively tested in the field -- there is no longer any circumstance in which archivists may lament an absence of direction in the area of personal archives.

Given the ever-changing nature of personal digital records, archivists must realize that there will never be a period of stasis, but only of relative calm when measures, no matter how imperfect they seem at the time, must be taken with the tools they have in order to capture the digital documentary heritage created by individual records creators.

The conclusion of this thesis collates and weighs the findings of each of its three chapters and proposes a hypothetical personal archiving strategy in the interest of continued development within this area of professional specialization.

CONCLUSION

REVISITING THE PRE-CUSTODIAL INTERVENTIONIST IDEAL

This thesis has argued significantly more upstream or pre-custodial effort must be invested in the archiving of personal digital records for the simple reason that more proactive measures are required in the capture and preservation of these materials than was previously the case with paper-based and analog documentary forms. Correspondingly, this thesis posits that the passivity of the traditional "end-of-lifecycle" approach to archiving personal records, though understandable in archiving analog documentary forms, is untenable in the digital age.

This thesis has introduced and discussed PIM perspectives on personal computing environments to reveal actualities of personal recordkeeping cultures, designations of value, and archiving strategies heretofore undisclosed by archival scholarship in the area of both personal and digital records. By leveraging the tacit knowledge of individuals in their day-to-day computing habits, PIM research seeks to improve human-computer interaction while at the same time communicates to the archivist information on the modes of personal records creation, prevailing types and classes of privately generated documentary forms, nuances in search and browse-based recordkeeping behaviours, the intricacies of appraisal decisions and private notions of value, as well as the complexities of digital preservation in both on and offline environments. Though PIM study is not the panacea for the difficulties encountered in the archiving of personal digital records, it nevertheless provides a conceptual framework through which archivists may come to better understand and document how personal records are created, accumulated, used, and preserved by individuals in the digital present. Just as research on the functions and

activities of contemporary governments has yielded new ways of archiving organizational electronic records, so too may research on pre-custodial personal recordkeeping behaviours and personal archiving strategies cultivate new and more effective means of archiving personal digital records.

The third chapter of this thesis has surveyed and assessed the current landscape of personal digital archives with a special emphasis on archival technology. As demonstrated in this chapter, archiving personal digital records is not simply a matter of dropping digital objects and metadata into a digital repository. Archiving personal records in a digital age involves expending a tremendous amount of effort in the rescue of data from fragile and decaying storage media, allocating a considerable amount of time and resources to both pre- and post-ingest processing activities, as well as a comprehensive knowledge of current and emerging technological standards, tools, and services. First-hand experience with the installation and configuration of software systems such as Archivematica and HOPPLA, populating multipurpose metadata schema, executing file transfers, generating hash check values, customizing digital migration tools, extracting metadata, or the processing of digital records from ingest to access in OAIS-based repositories are not yet required of archivists working primarily with paper-based personal records. However, the day is fast approaching when the predominance of digital documentary forms will make more than a general knowledge of digital tools and procedures, similar to those reviewed in this third chapter, a core competency alongside archival appraisal, arrangement, and description. In short, if their work must be divided between personal records creating and archival environments, can archivists interface

with pre-custodial streams of personal records mediation and if so, what would a possible strategy for personal digital archives look like today?

In the first chapter of this thesis, a number of unique theoretical perspectives on personal archives were discussed. Chief among them was the revolutionary pre-custodial interventionist approach proposed by Adrian Cunningham who believed in order to preserve and provide access to personal electronic records, archivists first needed to intervene in the earlier stages of records creation, management, and documentation. Somewhat hindered by the archival attitudes and technologies of the time, pre-custodial intervention seemed too "blue-sky" and intrusive to be applied in archival practice. Yet, a number of projects, approaches, and technologies reviewed in this thesis suggest pre-custodial intervention may be applied to real-world archival scenarios given the current technological climate. The final pages of this thesis present an illustration of a pre-custodial model for personal digital archives.

Pre-Custodial Intervention Strategy

Phase One – Initial and Recurrent Contact Activities[1]

In the event that archivists are able to identify individual records creators whose lives, careers, and hence archives will be of potential cultural or historical significance, there are a number of preliminary actions that may be performed in the pre-custodial environment. First, initial contact is made with the records creator at which time the frequency of future contact is discussed and agreed upon. In this first phase, the archivist surveys and documents local hardware and software onsite, performs a quantitative (physical volume) and qualitative (data types) assessment of obsolete/retired media, documents the make and model of mobile and other peripheral devices, and takes

[1] Phase One activity workflow is graphically represented in Appendix A.

photographic or video snapshots of the personal records creating environment(s). The archivist then interviews the creator to discover the number of online service providers (email, file sharing, and social media platforms) they are currently using and discusses scheduling for data synchronization between their personal computers, external hard drives, mobile devices, and cloud storage. Here, the creator may be asked about the different values their digital records hold to develop a personalized value folksonomy unique to that individual to assist in the eventual appraisal, intellectual arrangement, and description of their records by the archivist. While still onsite, the archivist confirms the Internet Protocol (IP) address of the creator's personal computer(s) to facilitate remote connection sessions and discusses the option of either periodic remote capture (performed by the archivist) or predetermined interval deposits (performed by creator).

Ideally, the archivist will remotely connect to the creator's personal computer as much as possible to avoid the known labour intensive and resource-taxing elements of onsite pre-custodial intervention.[2] The frequency and duration of these remote access sessions and the degree to which archivists may actually examine the content of the personal computer (such as restrictions on specific folder directories and hard drives) must also be agreed on upfront. In the interest of establishing records provenance and integrity control, the archivist may also wish to generate hash values (MD5 and SHA-1) for files and folders (or entire disk drives) which may then be recorded in a database registry. During this and successive contact sessions, it is advisable to capture information about the directory structures of the creator's personal computer(s) in the

[2] As described by Canadian archivist Lucie Paquet in "Appraisal, Acquisition and Control of Personal Electronic Records: From Myth to Reality," *Archives & Manuscripts* 28 (November 2000), pp. 71-91.

form of text files or screenshots as this records the initial original order of digital records as well as its evolution over time.

The archivist may install open-source software such as a suite of tools and services for online data repatriation and a drag-and-drop file transfer application with functionality and user-friendliness comparable to the FileZilla SFTP client. A more robust transfer program could also be developed to create a "digital drop-zone" where the creator drops digital records into a desktop icon which in turn automatically uploads the data to the archival server.[3] Installing a network-enabled home archiving system (such as HOPPLA) is also advisable for the management of records the creator has chosen to preserve according to their personal notions of value. Before concluding the onsite visit, the archivist should consider discussing the importance of performing regular full or incremental data backups, repatriations of data, as well as system diagnostics and virus checks and, whenever possible, arrange to have these tasks performed automatically through system configurations.[4]

Phase Two – Digital Acquisition Activities[5]

The second phase of the pre-custodial intervention model involves the actual acquisition of digital records, which ideally occurs at regular pre-determined intervals as opposed to a one-time accession. In the early stages of this second phase, attention should be directed toward at-risk data (such as that stored on floppy disks and ancestral computing platforms) first as these records may not exist or be accessible at the next

[3] To ensure the integrity of the records after transfer, a specification such as the Library of Congress BagIT program may be used. With BagIT, content is packaged along with a simple text-file manifest containing an inventory of files and a checksum for each file.
[4] Though it is unlikely to occur with any consistency, the archivist may ask the donor to add semantic metadata to their records (i.e. in the properties field of files) whenever possible.
[5] Phase Two activity workflow is graphically represented in Appendix B.

accession interval. Next, drawing on PIM insights, the archivist performs an initial appraisal of the creator's personal computer(s) to discover general recordkeeping behaviours (search or browse based tendencies), and to examine directory and folder structuring (frequent-filer or non-filer). Here, an appraisal of records and metadata within the already installed home archiving system should be considered to ensure near obsolete file formats have been migrated according to pre-programmed normalization paths that ideally have been automatically updated by the system's update service informed by archival digital preservation standards.

In acquiring digital records, the archivist may choose to prompt the donor to upload records to a secure archival server (via SFTP or a customized file transfer application) or choose to capture digital records through a disk image of specific folder directories or the entire personal computer system itself. Online data such as websites, blogs, or social media posts may also be captured at this point if the donor has actively repatriated and stored this content or, alternately, the archivist may harvest this online content providing password permissions have been obtained. At this stage, provenancial information created and accumulated over time within domains of personal computing has effectively reached its zenith for those records about to be accessioned and efforts should be made to ensure this valuable information is captured and represented in basic, archivist-authored, structured metadata. All digital objects and metadata acquired in this phase are then transferred to a stable offline digital environment for further archival processing.

Once the archivist has secured both physical and intellectual control of the personal digital records, the bit stream data captured through disk imaging processes

should be stored in a monitored preservation environment while surrogate master versions are prepared for ingest to a OAIS-based digital repository. Final pre-ingest procedures of this pre-custodial model include virus checks, corroboration of provenance and authenticity using the Phase One hash value registry, indexing of files and folder directories, augmentation of creator generated metadata (additional technical or contextual metadata), and the collation of all digital objects and their associated metadata. Before OAIS-ingest occurs, the archivist should perform both a technical and content appraisal of the records. Here, digital records are examined to ensure data has not been corrupted or otherwise lost through the comparison of hash values. Finally, the archivist must appraise the informational content of files, or groups of files, to separate the wheat from the chaff and find relevant and purposeful data. However, manual file-by-file appraisal is simply unreasonable considering the sheer volume of digital records captured during acquisitions. To this end, the archivist may employ fuzzy-hash techniques as a viable appraisal tool.

Though homologous digital files are not exactly identical, they do however contain identical sets of bytes in the same order. Fuzzy hashing is able to determine the relatedness of digital objects by calculating their hash values to identify files containing a high percentage of similarities. Using this technique, the archivist could hone in on and manually review only those text, image, video or web files with the highest percentage of similarities.[6] The ability to locate related digital records means archivists may trace the evolution of a textual work or video project, or select the most representative digital images from multiple, yet very similar versions. Given the abundance and duplication of

[6] Jesse Kornblum, "Identifying almost identical files using context triggered piecewise hashing," *Digital Investigation* 3 (2006), pp. 91-97. Available at http://www.sciencedirect.com/science/article/pii/S1742287606000764 (accessed 11 November 2011).

digital materials, fuzzy hashing should figure prominently in any appraisal of digital archives. Following these technical and content appraisals, the archivist prepares Submission Information Packages (SIPS) for OAIS-based repository ingest where the digital records receive additional processing in AIP and DIP stages including virus checks, metadata extractions, file formant identifications and validations, and format migrations for both preservation and access copies.

The Future of Archiving the Personal Digital Past

Even the best intentioned archival documentation strategy will overlook, either by design or omission, individuals in society whose digital records warrant long-term preservation by archival institutions. Acknowledging this reality, archivists cannot possibly intervene in the pre-custodial creation environments of people they have not yet identified. With this in mind, archivists may still continue to be proactive in the archiving of personal digital records through research aimed at gaining a better understanding of the personal pre-custodial environment while not necessarily directly intervening in its streams of personal records mediation. This pre-custodial understanding may be achieved by engaging in studies similar to those performed by the discipline of Personal Information Management or similar to the surveys and interviews carried out in the Paradigm and Digital Lives research projects. Case studies on collections of personal digital records, such as the Salman Rushdie and Michael Joyce archives, should also continue to be published in the interest of contributing to and eventually establishing proven best practices for the processing of personal digital archives. Continued research by archivists in the computing environments of individuals may also yield archivally inspired hardware and software programs or realistic metadata application profiles

tailored for individual records creators for, as seen in the case of Archivematica, archivists are now themselves capable of developing technologies to assist them in their work.[7] Once relegated to the margins of archival theory and practice, the specialized area of personal archives has gained considerable professional ground to become a major area of interest to which entire conferences, and most recently entire books, have been dedicated.[8]

In their discussion of the changing relationship between archival institutions and an increasingly digital society, information technology specialists Richard Katz and Paul Gandel note:

> The archivist cannot likely remain a creature exclusively of the tower. The values we share and the standards that we must promote, and the rich contextual knowledge about records in which we specialize, must be instantiated when and where the future historical record is being created, as well as into the culture of those technology providers whose products are reshaping the landscape of shared human memory.[9]

It is easy to become overwhelmed by the rapid evolution of digital technologies and the equally rapid change it brings to the technical processes of personal records creation, management, and preservation. One might say accelerated evolution has become an endemic quality of personal digital archives. And yet, for all of the developments in PIM tools, storage media, cloud computing, digital forensics, metadata, and digital repository architectures, there remain two constant and stabilizing elements of personal archives to consider. The first of these elements is the need of individuals to create records for

---

[7] Peter Van Garderen, the president of Artefactual Systems Inc. is a graduate of the University of British Columbia's Master of Archival Studies program.

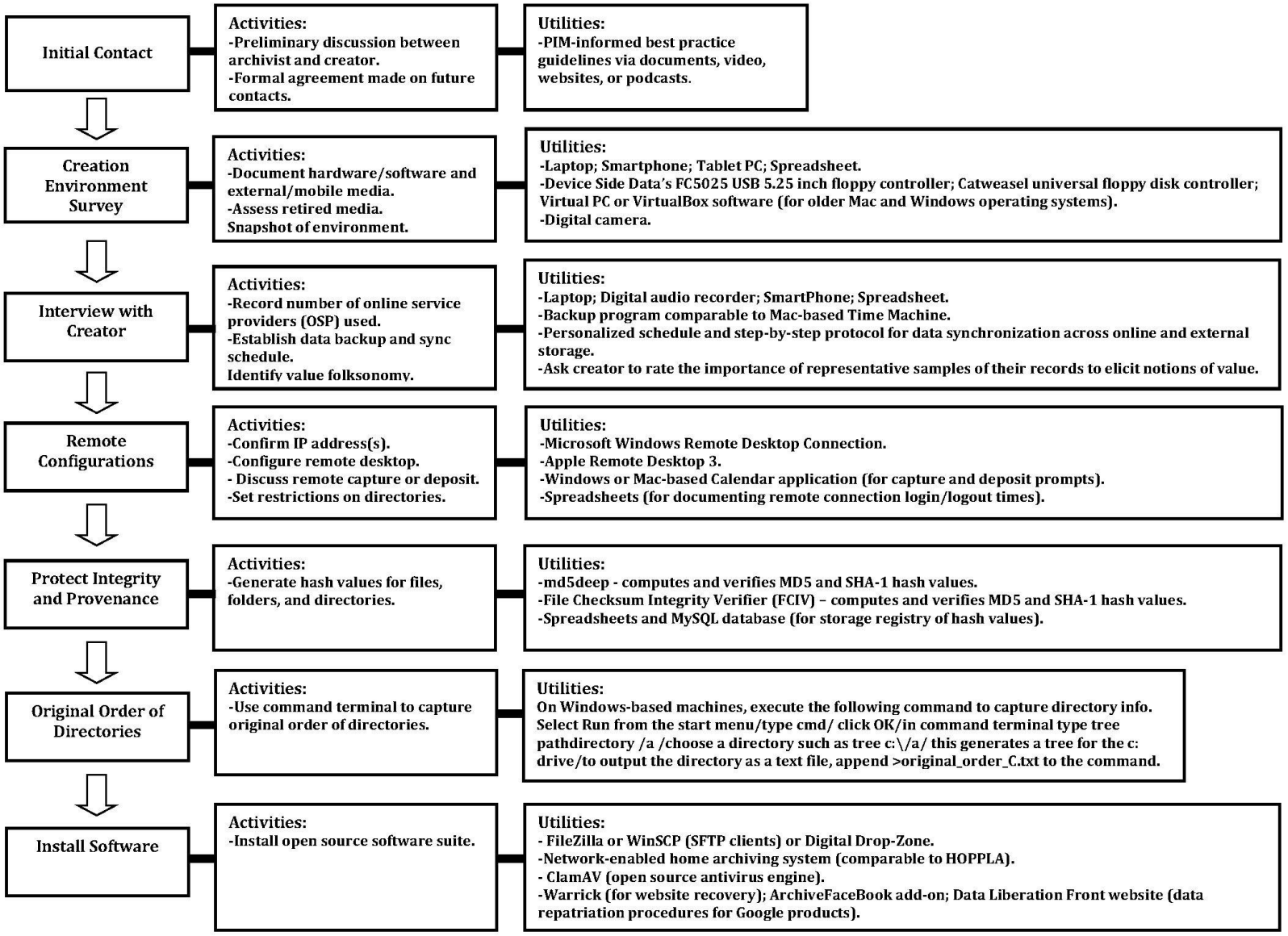
[8] For example the Personal Digital Archiving Conference, available at http://www.personalarchiving.com/ (accessed 22 November 2011), Richard Cox, *Personal Archives and a New Archival Calling: Readings, Reflections, and Ruminations*, (Duluth: Litwin Books, 2008), and Christopher A. Lee (ed.), *I, Digital: Personal Collections in the Digital Era*, (Chicago: Society of American Archivists, 2011).

[9] Richard N. Katz and Paul B. Gandel, "The Tower, the Cloud, and Posterity," in *Controlling the Past: Documenting Society and Institutions, Essays in Honor of Helen Willa Samuels*, Terry Cook, ed., (Chicago: Society of American Archivists, 2011), p. 236.

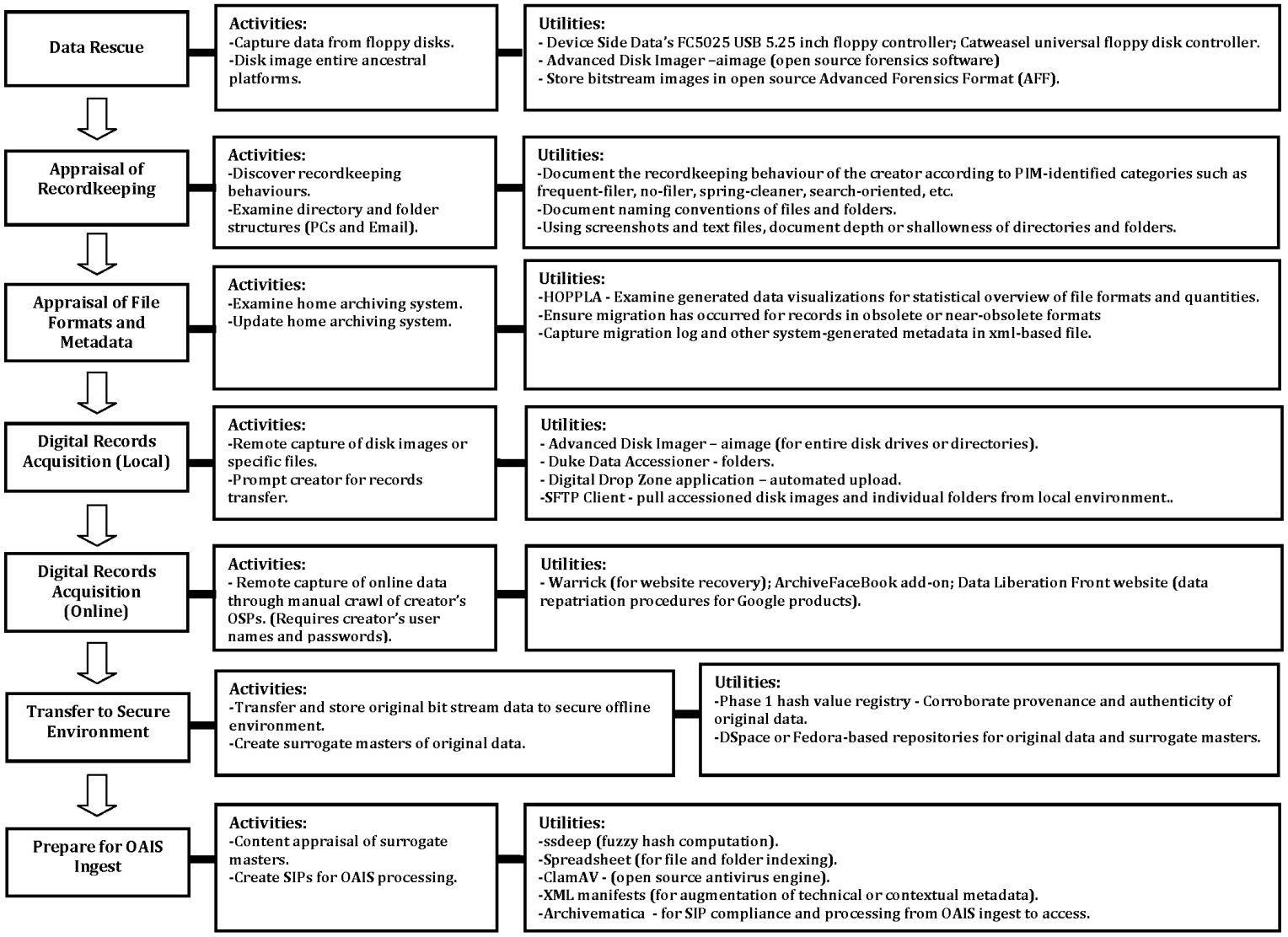
everyday reference, to sustain personal memory, to document interactions within society, to transmit knowledge from one generation to the next, and to construct and maintain self-identity.[10] In other words, while the medium for recording personal information is in a constant state of change, the motivations behind records creation are by comparison to technology quite stable and should be the primary focus of the archivist. The second stabilizing element of personal archives relates to value. The value they have for the people who create them and their value to collective documentary heritage are also largely similar regardless of the means of communication used, except that now the value of personal digital archives is more often deeply embedded in their binary code.

[10] Elisabeth Kaplan, "We Are What We Collect, We Collect What We Are: Archives and the Construction of Identity," *American Archivist* 63 (Spring/Summer 2000), pp. 126-151.

Appendix A: Pre-Custodial Workflow – Phase One



Appendix B: Pre-Custodial Workflow – Phase Two



BIBLIOGRAPHY

Websites:

Advanced Disk Imager (aimage). Available at http://afflib.org/software/aimage-the-advanced-disk-imager Accessed 24 November 2011.

Archivematica. Available at http://archivematica.org Accessed 24 November 2011.

Catweasel. Available at http://www.jschoenfeld.com/products/catweasel_e.htm Accessed 24 November 2011.

Data Liberation Front. Available at http://www.dataliberation.org/ Accessed 24 November 2011.

Digital Lives Research Project. Available at http://www.bl.uk/digital-lives/ Accessed 24 November 2011.

Device Side Data. Available at http://www.deviceside.com/ Accessed 24 November 2011.

DOS Games Discussion. Available at http://www.dosgames.com/forum/ Accessed 24 November 2011.

Dublin Core Metadata Initiative. Available at http://dublincore.org/

DropBox. Available at http://www.dropbox.com/ Accessed 24 November 2011.

Duke University Archives. Available at http://library.duke.edu/uarchives/ Accessed 24 November 2011.

Emory University Libraries – Manuscript, Archives, & Rare Book Library (MARBL). Available at http://marbl.library.emory.edu/ Accessed 24 November 2011.

Forensics Wiki. Available at http://www.forensicswiki.org/ Accessed 24 November 2011.

Home and Office Painless Persistent Long-term Archiving (HOPPLA). Available at http://www.ifs.tuwien.ac.at/dp/hoppla/ Accessed 24 November 2011.

InterPARES Project. Available at http://www.interpares.org/ Accessed 24 November 2011.

Library and Archives Canada (LAC). Available at http://www.collectionscanada.gc.ca Accessed 24 November 2011.

Library of Congress. Available at http://www.loc.gov Accessed 24 November 2011.

Microsoft Corporation MSDN Library. Available at http://msdn.microsoft.com/en-us/library Accessed 24 November 2011.

Mozilla. Available at http://www.mozilla.org/ Accessed 24 November 2011.

Personal Archives Accessible in Digital Media (Paradigm) Project. Available at http://www.paradigm.ac.uk/ Accessed 24 November 2011.

Society of American Archivists (SAA) Glossary of Archival and Records Terminology. Available at http://www.archivists.org/glossary/ Accessed 24 November 2011.

Ssdeep. Available at http://ssdeep.sourceforge.net/ Accessed 24 November 2011

University of Manitoba Archives and Special Collections. Available at http://umanitoba.ca/libraries/units/archives/ Accessed 24 November 2011.

Warrick. Available at http://warrick.cs.odu.edu/ Accessed 24 November 2011.

Books and Theses:

Bearman, David. *Electronic Evidence: Strategies for Managing Records in Contemporary Organizations*. Pittsburgh: Archives & Museum Informatics, 1994.

Boardman, Richard. "Improving Tool Support for Personal Information Management." Department of Electrical and Electronic Engineering, Imperial College London, PhD Dissertation, 2004.

Bodleian Library, John Rylands University Library of Manchester, and Joint Information Systems Committee, *Paradigm: Workbook on Personal Digital Archives*. Oxford: Bodleian Library, 2007.

Consultative Committee for Space Data Systems, *Reference Model for an Open Archival Information System (OAIS) Blue Book*. (January 2002).

Cox, Richard. *Personal Archives and a New Archival Calling: Readings, Reflections, and Ruminations*. Duluth: Litwin Books, 2008.

Ellis, Roger H and Peter Walne. eds. *Selected Writings of Sir Hilary Jenkinson*. Gloucester, 1980.

Jenkinson, Sir Hilary. *A Manual of Archival Administration*, new and revised edition. London: Percy, Lund, Humphries & Co Ltd, 1937.

John, Jeremy Leighton, Ian Rowlands, Peter Williams and Katrina Dean. *Digital Lives, Personal Digital Archives for the 21st Century: An Initial Synthesis, Beta Version 0.2*. (March 2010). Available at http://britishlibrary.typepad.co.uk/files/digital-lives-synthesis02-1.pdf

Jones, William and Jaime Teevan (eds.), *Personal Information Management*. Seattle: University of Washington Press, 2007.

Kirschenbaum, Matthew G. *Mechanisms: New Media and the Forensic Imagination*, Cambridge, Massachusetts: The MIT Press, 2008.

Lee, Christopher A. "Defining Digital Preservation Work: A Case Study of the Development of the Reference Model for an Open Archival Information System." University of Michigan School of Information, PhD Dissertation, 2005.

---. (ed.), *I, Digital: Personal Collections in the Digital Era*. Chicago: Society of American Archivists, 2011.

Muller, Samuel, Johan Feith, and R. Fruin. *Manual for the Arrangement and Description of Archives* (1898), 2d ed. (1940) Arthur H. Leavitt, trans. New York, reissued 1968.

National Information Standards Organization (NISO), *Understanding Metadata*, Maryland: NISO Press, 2004, Available at www.niso.org/publications/press/UnderstandingMetadata.pdf (accessed 15 May 2011).

Schellenberg, T.R. *Modern Archives: Principles and Techniques*. Chicago: University of Chicago Press, 1956.

Simon, Herbert. *Computers, Communications and the Public Interest*, Martin Greenberger, ed. Baltimore: Johns Hopkins University Press, 1971.

Taylor, Karyn. "From Paper to Cyberspace: Changing Communication Technologies and the Implications for Personal Records Archivists," University of Manitoba, Department of History (Archival Studies) MA Thesis (2002).

Zhang, Jane. "The Principle of Original Order & The Organization and Representation of Digital Archives," Faculty of the Simmons College Graduate School of Library and Information Science, PhD Dissertation, (2010).

Articles, Chapters, Reports, and Proceedings:

Abrams, David, Ron Baecker, and Mark Chignell. "Information archiving with bookmarks: personal Web space construction and organization," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, California: ACM Press, 1998, pp. 41-48.

Alberts, Inge, Jen Schellinck, Craig Eby, and Yves Marleau. "Bridging Functions and Process for Records Management," *Canadian Journal of Information and Library Science*, 34:4 (December 2010), pp. 365-690.

Armstrong, Pam. "Becoming a Trusted Digital Repository, Library and Archives Canada: Managing Information in the Public Sector, Meeting the Challenge," *Ontario Information Management Conference*, Toronto, Ontario, April 28, 2009, available at www.verney.ca/opsim2009/presentations/783.pdf (accessed 09 August 2011).

Atherton, Jay. "From Life Cycle to Continuum: Some Thoughts on the Records Management – Archives Relationship," *Archivaria* 21 (Winter 1985-86), pp. 43-51.

Bak, Greg and Pam Armstrong. "Points of convergence: seamless long-term access to digital publications and archival records at Library and Archives Canada," *Archival Science* 8:4 (2008), pp. 279-293.

Balter, Olle. "Strategies for organizing email," in *People and Computers XII: Proceedings of HCI'97*, H. Thimbley, B. O'Conaill and P.J. Thomas, eds. London: Springer-Verlag, 1997, pp. 21-38.

Barreau, Deborah and Bonnie A. Nardi. "Finding and Reminding: File Organization from the Desktop," *ACM SIGCHI Bulletin* 27:3 (July 1995), pp. 39-43.

Barton, Jane, Sarah Currier, and Jessie M.N. Hey. "Building Quality Assurance into Metadata Creation: an Analysis based on the Learning Objects and e-Prints Communities of Practice," in Sutton, S. and Greenberg, J. and Tennis, J., eds. *Proceedings 2003 Dublin Core Conference: Supporting Communities of Discourse and Practice - Metadata Research and Applications* Seattle: Information Institute of Syracuse, 2003. Available at http://eprints.erpanet.org/83/ (accessed 25 October 2011).

Beagrie, Neil. "Plenty of Room at the Bottom? Personal Digital Libraries and Collections," *D-Lib Magazine*, 11:6 (June 2005). Available at http://www.dlib.org/dlib/june05/beagrie/06beagrie.html (accessed 2 June 2011).

Bearman, David. "An Indefensible Bastion: Archives as a Repository in the Electronic Age," in *Archives and Museum Informatics Technical Report 13 (1991)*, David Bearman, ed. Pittsburgh: Archives & Museum Informatics, 1991, pp. 14-24.

Bell, Gordon and Jim Gemmell. "A Digital Life," *Scientific American* 296 (March 2007), pp. 58-65.

Boardman, Richard, and Martina Angela Sasse. "Stuff Goes into the Computer and Doesn't Come Out: A Cross-tool Study of Personal Information Management," in *Proceedings of the SIGCHI conference on Human factors in computing systems, CHI*, 6:1(New York: ACM Press, 2004), pp. 583-590.

Hans Booms. "Society and the Formation of a Documentary Heritage: Issues in the Appraisal of Archival Sources," *Archivaria* 24 (Summer 1987), pp. 69-107.

---."Überlieferungsbildung: Keeping Archives as a Social and Political Activity," *Archivaria* 33 (Winter 1991-92), pp. 25-33.

Canadian Association of University Teachers (CAUT). "Open Letter to Daniel Caron, Librarian and Archivist of Canada," (June 2011). Available at http://www.caut.ca/uploads/CAUT_to_Caron_LACv2.pdf (accessed 17 October 2011).

Caplan, Pricilla. *Understanding PREMIS*, (Library of Congress, 2009), pp. 1-26.

Caron, Daniel J. and Richard Brown. "The Documentary Moment in the Digital Age: Establishing New Value Propositions for Public Memory," *Archivaria* 71 (Spring 2011), pp. 1-20.

Carr, Leslie and Stevan Harnard. "Keystroke Economy: A Study of the Time and Effort Involved in Self-Archiving," (University of Southampton, 2005) available at http://eprints.ecs.soton.ac.uk/10688/ (accessed 25 September 2011)

Carroll, Laura L. "Arranging and Describing Born-digital Archives: The Salman Rushdie Papers at Emory University," PowerPoint Presentation, *Society of American Archivists Conference*, Chicago, Illinois, August 26, 2011, available at http://e-records.chrisprom.com/wp-content/uploads/2011/08/PAERWhatWorksNow.pdf (accessed 25 August 2011)

---. "The Salman Rushdie Papers at Emory University: Processing a Born Digital Manuscript Collection," PowerPoint Presentation, *Association of Canadian Archivists Institute on Personal Archives*, Toronto, Ontario, October 29 and 30, 2010.

Carroll, Laura L and Erika Farr, "Hybrid teams for hybrid archives: Collaboration and born-digital archives," PowerPoint Presentation, *Association of College and Research Libraries: Rare Books and Manuscripts Section (RBMS) Conference*, Philadelphia, Pennsylvania, June 22 and 25, 2010.

Chervenak, Ann L, Vivekanand Vellanki and Zachary Kurmas. "Protecting File Systems: A Survey of Backup Techniques," *Proceedings of the Joint IEEE and NASA Mass Storage Conference*, (March 1998), pp. 17-32.

Coffin, Robert. "Information Management and e-Government in Canada: From Government On Line to Service Transformation," Online presentation available at http://www.docstoc.com/docs/71273777/Developing-GOL-Funding-Options (accessed 24 August 2011).

Cook, Terry. "An Archival Revolution: W. Kaye Lamb and the Transformation of the Archival Profession," *Archivaria* 60 (Fall 2005), pp. 185-234.

Cook, Terry, ed. "Documenting Society and Institutions: The Influence of Helen Willa Samuels," in *Controlling the Past: Documenting Society and Institutions, Essays in Honor of Helen Willa Samuels,* Chicago: Society of American Archivists, 2011, pp. 1-28.

----."Electronic Records, Paper Minds: The revolution in information management and archives in the post-custodial and post-modernist era," *Archives & Manuscripts* 22 (November 1994), pp. 300-328.

---."Leaving Archival Electronic Records in Institutions: Policy and Monitoring Arrangements at the National Archives of Canada," *Archives and Museum Informatics* 9:2 (1995), pp. 141-149.

---. "Macro-appraisal and functional analysis: documenting governance rather than government," *Journal of the Society of Archivists* 25:1 (2004), pp. 5-18.

---."Mind Over Matter: Towards a New Theory of Archival Appraisal," in *The Archival Imagination: Essays in Honour of Hugh A. Taylor*, Barbara L. Craig, ed. Ottawa: Association of Canadian Archivists, 1992, pp. 38-70.

---."The Impact of David Bearman on Modern Archival Thinking: An Essay of Personal Reflection and Critique," *Archives and Museum Informatics* 11:1 (1997), pp. 15-37.

Cox, Richard J. "The Record in the Manuscript Collection," *Archives & Manuscripts* 24:1 (May 1996), pp. 52-61.

Cunningham, Adrian. "Archival Institutions," in *Archives: Recordkeeping in Society*, Sue McKemmish, Michael Piggott, Barbara Reed and Frank Upward, eds. New South Wales: Centre for Information Studies, Charles Sturt University, 2005, pp. 21-50.

---. "Beyond the Pale? The 'flinty' relationship between archivists who collect the private records of individuals and the rest of the archival profession," *Archives & Manuscripts* 24 (May 1996), pp. 20-26.

---."The Archival Management of Personal Records in Electronic Form: Some Suggestions," *Archives & Manuscripts* 22 (May 1994), pp. 94-105.

---."Waiting for the Ghost Train: Strategies for Managing Personal Records Before it is Too Late," *Archival Issues* 24 (1999), pp. 55-64.

Davenport, Thomas H. and John C. Beck. "The Attention Economy," *Ubiquity* (May 2001). Available at http://ubiquity.acm.org/article.cfm?id=376626 (accessed 10 October 2011).

Davis, Susan E. "Electronic Records Planning in 'Collecting Repositories,' *The American Archivist* 71 (Spring/Summer 2008), pp.167-189.

Ducheneaut, Nicholas and Victoria Bellotti. "Email as a habitat: An exploration of embedded personal information management," *ACM Interactions* 8:5 (September/October 2001), pp.30-38.

Duranti, Luciana. "Diplomatics: New Uses for an Old Science," *Archivaria* 28 (Summer 1989), pp. 7-27.

---. "Archives as a Place," *Archives & Manuscripts* 24 (1996), pp. 242-255.

Eastwood, Terry. "What is Archival Theory and Why is it Important?" *Archivaria* 37 (Spring 1994), pp. 122-130.

Etherton, Judith. "The role of archives in the perception of the self," *Journal of the Society of Archivists* 27:2 (2006), pp. 227-246.

Farr, Erika. "When Papers Aren't Just Paper: Hybrid Archives at Emory," PowerPoint Presentation, *Digital Lives Conference*, London, England, February 9, 2009.

Fertig, Scott, Eric Freeman, and David Gelernter. "Finding and Reminding Reconsidered," *ACM SIGCHI Bulletin* 28:1 (January 1996), pp. 66-69

Featherstone, Mike. "Archive," *Theory, Culture and Society*, 23 (2006), pp. 591-596.

Fisher, Danyel, A.J. Brush, Eric Gleave, and Marc A. Smith. "Revisiting Whittaker & Sidner's 'Email Overload' Ten Year Later," in *Proceedings of the 15th International Conference on Intelligent User Interfaces,* (November 2006), pp. 309-312.

Fisher, Robert. "In Search of a Theory of Private Archives: The Foundational Writings of Jenkinson and Schellenberg Revisited," *Archivaria* 67 (Spring 2009), pp.1-24.

---. "The Grandmother's Story: Oral Tradition, Family Memory, and a Mysterious Manuscript," *Archivaria* 57 (Spring 2004), pp. 107-130.

Forstrom, Michael. "Managing Electronic Records in Manuscript Collections: A Case Study from the Beinecke Rare Book and Manuscript Library," *The American Archivist* 72 (Fall/Winter 2009), pp. 460-477.

Gilliland-Swetland, Luke J. "The Provenance of a Profession: The Permanence of the Public Archives and Historical Manuscripts Traditions in American Archival Theory," *The American Archivist* 54 (Spring 1991), pp. 160-175.

Goldman, Ben. "Bridging the Gap: Taking Practical Steps Toward Managing Born-Digital Collections in Manuscript Repositories," *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage* 12:1 (Spring 2011), pp. 11-24.

Government of Canada. *Libraries and Archives Canada, 2010-2011, Report on Plans and Priorities*. Available at http://www.collectionscanada.gc.ca/012/012-205-e.html (accessed 17 October 2011).

---. "*Libraries and Archives Canada, 2010-2011, Report on Plans and Priorities*," pp.1-34. Available at http://www.collectionscanada.gc.ca/012/012-205-e.html (accessed 24 August 2011).

---. "*Records, Document and Information Management System (RDIMS) User Guide*. Available at http://www.oic-ci.gc.ca/eng/rr-sl-odi-adi.aspx (accessed 25 December 2011).

Greenberg, Jane et al. *Final Report for the AMeGA (Automatic Metadata Generation Applications) Project*, Chapel Hill: University of North Carolina School of Information and Library Science, 2005.

Greene, Mark A. "The Power of Meaning: The Archival Mission in the Postmodern Age," *The American Archivist* 65 (Spring/Summer 2002), pp. 42-55.

Gwizdka, Jacek. "Email Task Management Styles: The Cleaners and the Keepers," in *SIGCHI Conference on Human Factors in Computing Systems*, Vienna: ACM Press, 2004, pp. 1235-1238.

Hackett, Yvette. "The Search for Authenticity in Electronic Records," *The Moving Image 3*, no.2 (2003), pp. 100-107.

Hackman, Larry J, and Joan Warnow-Blewett, "The Documentation Strategy Process: A Model and a Case Study," *The American Archivist* 50 (Winter 1987), pp. 12-47.

Ham, F. Gerald. “Archival Strategies for the Post-Custodial Era,” *The American Archivist* 44 (Summer 1981), pp. 207-216.

---.“The Archival Edge,” *The American Archivist* 38 (January 1975), pp. 5-13.

Harris, Verne. “On the Back of a Tiger: Deconstructive Possibilities in 'Evidence of Me',” *Archives & Manuscripts* 29:1 (May 2001), pp. 8-21.

Heery, Rachel and Manjula Patel. “Application profiles: mixing and matching metadata schemas,” *Ariadne* 25 (September 2000). Available at http://www.ariadne.ac.uk/issue25/app-profiles/ (accessed 15 May 2011).

Henderson, Sarah. “Personal Document Management Strategies,” *CHINZ '09 Proceedings of the 10th International Conference NZ Chapter of the ACM's Special Interest Group on Human- Computer Interaction*, (2009), pp. 69-76.

Hirsch, Rebecca. “The Permanence of Provenance: The 'Two Traditions' and the American Archival Profession,” *Journal of Archival Organization* 8:1 (2010), pp. 54-72.

Hives, Christopher L. “Thinking Globally, Acting Locally,” *Archivaria* 38 (Fall 1994), pp. 157-161.

Hobbs, Catherine. “Reenvisioning the Personal: Reframing Traces of Individual Life,” in *Currents of Archival Thinking*, Terry Eastwood and Heather MacNeil, eds. Santa Barbara: Libraries Unlimited, 2010, pp. 213-241.

---. “The Character of Personal Archives: Reflections on the Value of Records of Individuals,” *Archivaria* 52 (Fall 2001), pp. 126-135.

Hodge, Gail M. “Best Practices for Digital Archiving: An Information Life Cycle Approach,” *D-Lib Magazine* 6:1 (January 2000). Available at http://www.dlib.org/dlib/january00/01hodge.html (accessed 25 October 2011).

John, Jeremy Leighton. “Adapting Existing Technologies for Digitally Archiving Personal Lives: Digital Forensics, Ancestral Computing, and Evolutionary Perspectives and Tools,” *iPRES 2008: The Fifth International Conference on Preservation of Digital Objects* (2008). Available at http://www.bl.uk/ipres2008/presentations_day1/09_John.pdf (accessed 25 October 2011).

Johnson, Elizabeth Snowden. “Our Archives, Our Selves: Documentation Strategy and the Re-Appraisal of Professional Identity,” *The American Archivist* 71 (Spring/Summer 2008), pp. 190-202.

Kaplan, Elizabeth. "We Are What We Collect, We Collect What We Are: Archives and the Construction of Identity," *The American Archivist* 63 (Spring/Summer 2000), pp. 126-151.

Katz, Richard N. and Paul B. Gandel. "The Tower, the Cloud, and Posterity," in *Controlling the Past: Documenting Society and Institutions, Essays in Honor of Helen Willa Samuels*, Terry Cook, ed. Chicago: Society of American Archivists, 2011, pp. 217-238.

Kaye, Joseph et al. "To Have and to Hold: Exploring the Personal Archive," *Communications of the ACM* 49:1 (2006), pp. 275-284.

Kiehne, Thomas, Vivian Spoliansky, and Catherine Stollar. "From Floppies to Repository: A Transition of Bits, A Case Study in Preserving the Michael Joyce Digital Papers at the Harry Ransom Center," Unpublished paper, May 2005, available at https://pacer.ischool.utexas.edu/handle/2081/941 (accessed 25 September 2011).

Kirk, David and Abigail Sellen. "On Human Remains: Excavating the Home Archive," *Microsoft Technical Report* (June 2008). Available at http://research.microsoft.com/apps/pubs/default.aspx?id=70595 (accessed 15 May 2011), pp. 1-10.

Kornblum, Jesse. "Identifying almost identical files using context triggered piecewise hashing," *Digital Investigation* 3 (2006), pp. 91-97. Available at http://www.sciencedirect.com/science/article/pii/S1742287606000764 (accessed 11 November 2011).

Kuny, Terry. "A Digital Dark Ages? Challenges in the Preservation of Electronic Information, in *Sixty-third IFLA Council and General Conference*, (August 1997), pp. 1-12.

Kwasnik, Barbara H. "How a personal document's intended use or purpose affects its classification in an office," in *Proceedings of the ACM-SIGIR 12th Annual International Conference on Research and Development in Information Retrieval* (June 1989), pp. 207-210.

Lansdale, M.W. "The psychology of personal information management," *Applied Ergonomics* 19:1 (March 1988), pp. 55-66.

Library and Archives Canada. *Acquisition Priorities: Private Archives Working Group, Final Report: Acquisition Priorities: Setting Priorities for Private Archives*. Available at http://www.collectionscanada.gc.ca/modernization/012004-2055.01-e.html (accessed 24 August 2011).

---. *Departmental Performance Report 2010-2011*. Available at http://publications.gc.ca/collections/collection_2011/collectionscanada/SB1-4-2011-eng.pdf (accessed 13 December 2011)

---. *Preliminary Survey of a System Under Development Audit of the AMICAN Catalytic Initiative*, (June 2007). Available at http://nlc-bnc.ca/obj/012014/f2/012014-205-e.pdf (accessed 24 August 2011).

Lipowicz, Alice. "White House attention nudges e-archive toward completion by September 2011," *Federal Computer Week* (August 26, 2010). Available at http://fcw.com/articles/2010/08/26/white-house-nudges-electronic-archives-to-completion-by-september-2011.aspx (accessed 24 August 2011)

Lofus, Mary J. "The Author's Desktop," *Emory Magazine* (Winter 2010) available at http://www.emory.edu/EMORY_MAGAZINE/2010/winter/authors.html (accessed 16 August 2011).

Mackay, Wendy E. "More than Just a Communication System: Diversity in the Use of Electronic Mail," in *Proceedings of the 1988 ACM conference on Computer-supported cooperative work*. New York: ACM Press, 1998, pp. 344-353.

MacNeil, Heather. "Providing Grounds for Trust: Developing Conceptual Requirements for the Long-Term Preservation of Authentic Electronic Records," *Archivaria* 54 (Fall 2002), pp. 52-78

Malone, Thomas W. "How Do People Organize Their Desks? Implications for the Design of Office Information Systems," *ACM Transactions on Office Information Systems*, 1:1 (January 1983), pp. 99-112.

Marshall, Catherine C. "Rethinking Personal Digital Archiving, Part 1: Four Challenges from the Field," *D-Lib Magazine* (March/April 2008). Available at http://www.dlib.org/dlib/march08/marshall/03marshall-pt1.html (accessed 2 June 2011).

---."Rethinking Personal Digital Archiving Part 2: Implications for Services," Applications, and Institutions," *D-Lib Magazine* (March/April 2008). Available at http://www.dlib.org/dlib/march08/marshall/03marshall-pt2.html (accessed 15 May 2011).

Marshall, Catherine C., Sara Bly, and Francoise Brun-Cottan. "The Long Term Fate of Our Digital Belongings: Towards a Service Model for Personal Archives," *Proceedings of Archiving 2006*, Available at http://research.microsoft.com/apps/pubs/default.aspx?id=75527 (accessed 15 May 2011).

Marshall, Jennifer A. "Documentation Strategies in the Twenty-First Century?: Rethinking Institutional Priorities and Professional Limitations," *Archival Issues* 23:1 (1998), pp. 59-74.

McCown, Frank and Michael L. Nelson. "What Happens when Facebook is Gone?" *Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries* (2009), available at http://portal.acm.org/citation.cfm?id=1555440 (accessed 15 July 2011).

McCown, Frank, Joan A. Smith, and Michael L. Nelson. "Lazy Preservation: Reconstructing Websites by Crawling the Crawlers," *Proceedings of ACM WIDM* (2006). Available at http://www.cs.odu.edu/~fmccown/pubs/lazyp-widm06.pdf (accessed 15 July 2011).

McDonald, Robert A.J. "Who is Preserving Private Records?" *Archivaria* 38 (Fall 1994), pp. 155-157.

---. "Acquiring and Preserving Private Records: Cultural versus Administrative Perspectives," *Archivaria* 38 (Fall 1994), pp. 162-163.

McKemmish, Sue. "Evidence of me…," *Archives & Manuscripts* 24:1 (May 1996), pp. 28-45.

Meehan, Jennifer. "Rethinking Original Order and Personal Records," *Archivaria* 70 (Fall 2010), pp. 27-44.

Mell, Peter and Timothy Grance. "The NIST Definition of Cloud Computing (Draft)," National Institute of Standards and Technology, U.S. Department of Commerce, (January 2011), available at www.nist.gov/itl/cloud/upload/cloud-def-v15.pdf (accessed 15 July 2011).

Millar, Laura. "Discharging our Debt: The Evolution of the Total Archives Concept in English Canada," *Archivaria* 46 (Fall 2001), pp. 103-146.

---. "Touchstones: Considering the Relationship between Memory and Archives," *Archivaria* 61 (Spring 2006), pp. 105-126.

Moore, Gordon E. "Cramming more components onto integrated circuits," *Electronics* 38:8 (April 1965), pp. 1-4.

Nahuet, Robert. "The Management of Textual Digital Archives: A Canadian Perspective Library and Archives Canada and Federal Government Institutions," *Atlanti*, 17:1-2 (2007), pp. 29-45.

Nesmith, Tom. "Reopening Archives: Bringing New Contextualities into Archival Theory and Practice," *Archivaria* 60 (Fall 2005), pp. 259-274.

Newman, Sally. “Aileen Palmer’s Textual Lives,” in *The Intimate Archive: Journeys Through Private Papers*, Stephanie Owen Reeder and Tina Mattei, eds. Canberra: National Library of Australia, 2009, pp. 133-176.

Paquet, Lucie. “Appraisal, Acquisition and Control of Personal Electronic Records: From Myth to Reality,” *Archives & Manuscripts* 28 (November 2000), pp. 71-91.

Reed, Barbara. “Reading the Records Continuum: Interpretations and Explorations,” Available at www.records.com.au/pdf/Reading_the_Records_Continuum.pdf (accessed 11 June 2011).

Rosenberg, Victor. “The power of a family archive,” *Archival Science* 11 (2011), pp. 77-93.

Rosenthal, David S.H. et al. “Requirements for Digital Preservation Systems: A Bottom-Up Approach,” *D-Lib Magazine* 11:11 (November 2005). Available at http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html (accessed 25 September 2011).

Rothenberg, Jeff. “Ensuring the Longevity of Digital Documents,” *Scientific American* 272:1 (January 1995), pp. 42-47.

Samuels, Helen. “Who Controls the Past?” *The American Archivist* 49 (Spring 1986), pp. 109-124.

Scerri, Simon, Siegfried Handschuh and Stefan Decker. “Semantic Email as a Communication Medium for the Social Semantic Desktop,” *Communications* 5021:4 (2008), pp. 124-138.

Schwartz, Joan M. “'We make our tools and our tools make us': Lessons from Photographs for the Practice, Politics, and Poetics of Diplomatics,” *Archivaria* 40 (Fall 1995), pp. 40-74.

Shirley Spragge. “The Abdication Crisis: Are Archivists Giving Up Their Cultural Responsibility?” *Archivaria* 40 (Fall 1995), pp. 173-181.

Smith, Johanna and Pam Armstrong. “Preserving the Digital Memory of the Government of Canada: Influence and Collaboration with Records Creators,” *iPRES 2009: the Sixth International Conference on Preservation of Digital Objects,* Mission Day Conference Center, San Francisco, California, California Digital Library, UC Office of the President, October 5-6, 2009, pp. 184-185. Available at www.cdlib.org/services/uc3/iPres/presentations/SmithArmstrong.pdf (access 24 August 2011).

Stapleton, Richard. "Jenkinson and Schellenberg: A Comparison," *Archivaria* 17 (Winter 1983- 84), pp. 75-85.

Stollar, Catherine. "When Not All Papers Are Paper: A Case Study in Digital Archivy," *Provenance* 24 (2006), pp. 23-35.

Stollar, Catherine and Thomas Kiehne. "Guarding the Guards: Archiving the Electronic Records of Hypertext Author Michael Joyce," *New Skills for a Digital Era*, Washington, D.C. May 31-June 2, 2006. Available at www.lib.az.us/diggovt/documents/pdf/4_Stollar_Kiehne.pdf (accessed 10 August 2011).

Storch, Susan E. "Diplomatics: Modern Archival Method or Medieval Artifact?" *The American Archivist* 61 (Fall 1998), pp. 365-383.

Strodl, Stephan et al. "Automating logical preservation for small institutions with Hoppla," available at http://www.ifs.tuwien.ac.at/dp/hoppla/ (accessed 25 September 2011).

Thomas, Susan and Janette Martin. "Using the Papers of Contemporary British Politicians as a Testbed for the Preservation of Digital Personal Archives," *Journal of the Society of Archivists* 27:1 (April 2006), pp. 29-56.

Thomas, Susan. *A Practical Approach to the Preservation of Personal Digital Archives: Final Report to the JISC*, Oxford: Bodleian Library, 2007, pp. 1-38.

Upward, Frank and Sue McKemmish. "Somewhere Beyond Custody," *Archives & Manuscripts* 22 (May 1994), pp. 137-149.

Williams, Pete, Katrina Dean, and Jeremy Leighton John. "Digital Lives: Report of Interviews with the Creators of Personal Digital Collections," *Ariadne* 55 (April 2008). Available at http://www.ariadne.ac.uk/issue55/williams-et-al/ (accessed 10 May 2011).

Wilson, Ian E. "A Noble Dream: The Origins of the Public Archives of Canada," *Archivaria* 15 (Winter 1982-83), pp. 16-35.

Whittaker, Steve, Victoria Bellotti, and Jacek Gwizdka. "Email in personal information management," *Communications of the ACM - Personal information management* 49:1 (January 2006), pp. 68-73.

Whittaker, Steve and Julia Hirschberg. "The character, value and management of personal archives," *ACM Transactions on Computer Human Interaction* 8 (2001), pp. 150-170.

Whittaker, Steve and Candace Sidner. "Email overload, exploring personal information management of email," *Proceedings of the Conference on Human Factors in Computing Systems (CHI),* (1996), pp. 276-283.

Wilkes, Wolfgang, et al. "Towards Support for Long-Term Digital Preservation in Product Lifecycle Management," *The International Journal of Digital Curation* 1:6 (2011), pp. 282-296.

Williams, Caroline. "Personal Papers: Perceptions and Practices," in *What are Archives? Cultural and Theoretical Perspectives: A Reader,* Louise Craven, ed. Aldershot, England: Ashgate Publishing Company, 2008, pp. 53-67.