Another Bit Bytes the Dust:

The Technological and Human Challenges of Digital Preservation

By Chris Zaste

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

MASTER OF ARTS

Department of History, (Archival Studies)
Joint Master's Program
University of Manitoba/University of Winnipeg
Winnipeg, Manitoba

Table of Contents

Abstract

Digital communication produces millions of emails, text messages, movies, images, and much more every day. As with all historical records, digital records are important to preserve because they allow us to study the past. There are, however, several challenges regarding their preservation. Unlike many of their analogue counterparts, digital records rely on a combination of hardware and software to be accessible, but hardware and software eventually degrade and become obsolete. This makes digital records inaccessible because the means to render them are no longer available. In addition to these technological challenges, there are issues surrounding appraisal, copyright, significant properties, and metadata.

This thesis studies the challenges of digital preservation and what is being done to address them. I begin by introducing the challenges surrounding this topic and the methods of preservation that are currently available to archivists. I then analyse leading digital preservation standards such as the Reference Model for an Open Archival Information System (OAIS) and Preservation Metadata: Implementation Strategies (PREMIS) as well as digital preservation systems including Archivematica and Preservica. I also conduct a case study of Archivematica to analyse how well it manages the challenges of digital preservation. I conclude by explaining that there are no perfect solutions to digital preservation problems. The best that can currently be done is to manage the issues rather than solve them.

Acknowledgments

First, I would like to thank everyone who has helped me on this journey. I would like to express my deepest gratitude to my advisor, Dr. Greg Bak, for his guidance and advice on my thesis. I would also like to think Dr. Tom Nesmith who also provided assistance and guidance throughout this process. Together, Tom and Greg have been invaluable in helping me complete this thesis. I would also like to thank all of the organizations that I have worked for during this process. The experience that I gained from my coworkers has helped me with writing this thesis. Special thanks goes to the Mennonite Heritage Centre Archives for allowing me to conduct a case study at their facility.

Last but not least, I would like to thank all the members of my family. My parents, Denis and Michelle, and my sisters, Danielle and Melissa, have provided me with understanding and support throughout this process. Thank you to everyone. Your help and support have meant a lot to me.

Introduction

The widespread use of digital technology has connected people around the world like never before. In our digital culture, communication is now easy and nearly instantaneous. The products of this communication, digital records, have therefore become commonplace. Texts, emails, movies, websites, books, and many other digital records are widely used, but they will all lose both valuable content and context without digital preservation. Unlike most analogue records, digital records require specific hardware and software in order to be rendered. If these are not available, the digital record will be inaccessible to users. Furthermore, even if a digital record has been kept secure and is accessible, it may lack valuable contextual information that would have given meaning to the record. Despite their centrality in our digital culture, digital records are not easily preserved for future generations.

In response to these challenges, archivists[1] have developed a wide variety of methods to mitigate the difficulties of preserving digital records. These challenges and methods of digital preservation are the subject of this thesis. In chapter one, I will highlight the various technological and human challenges of digital preservation. Chapter two will analyze the responses of archivists to the challenges described in chapter one. This will include a detailed evaluation of the Open Archival Information System (OAIS) and Preservation Metadata: Implementation Strategies (PREMIS). It will also analyze preservation systems with a particular focus on Artefactual's Archivematica and Tessella's Preservica. Chapter three consists of a case study of an implementation of Archivematica by the author at the Mennonite Heritage Centre Archives in Winnipeg, as well as my own experimentations with it. Finally, the conclusion will

[1] Many professions are involved in digital preservation. These include archivists, librarians, record managers, computer scientists, and several others. Each of these fields contributes to both the academic and non-academic literature on digital preservation. This thesis uses writings from all of these fields and I am choosing the term archivist to denote the role that these professions play in digital preservation.

consist of a consideration of the effectiveness of the methods of digital preservation in mitigating the preservation challenges that I described in the introduction. I hope that this thesis will help illuminate the current state of the digital preservation field.

I also hope that despite the technical nature of digital preservation, we do not forget the human behind the machine. Digital records are created by people for their own purposes and this fact should not be overlooked by archivists. The archivist needs to select what is worth preserving, how it should be preserved, how much metadata to add, as well as navigating the financial and institutional challenges of managing an archive. There is no single coprrect response to these challenges. This inevitably leads to an archivist exercising his or her own judgement and human judgment is subjective. As a result, archivists can and do adopt different approaches to the challenges of digital preservation. I introduce this subjective human element in Chapter One and I hope they are studied more in the future.

Before any in-depth analysis of digital preservation can take place, it is important to understand the reasons why so many archivists are devoting significant time and effort to digital preservation. Without the intervention of archivists, there is a high probability that digital records will be lost to future generations. With so few traces of our digital lives and activities, there will be little evidence of past actions for people to keep and study. The few digital records that may survive the passage of time without the intervention of archivists will lack contextual information, which will render them unintelligible for future generations. These records will also be without proper metadata that describes their provenance and technical features. Without digital preservation, records will either be lost completely or lose essential contextual information which will render them unintelligible.

Failure to preserve digital records will result in negative effects on the various users of archives. The archival record can be used for medical research and to study climate conditions. For example the archives of the School Sisters of Notre Dame were an integral source of information for researchers studying the effects of dementia and Alzheimer's disease on the elderly.[2] In her thesis, *Medical Records Redefined: The Value of the Archival Record in Medical Research*, Natalie Vielfaure examines additional uses of archival medical records.[3] In his thesis, *Archives, Historical Climate Records, and the Climate Observations of Thomas Corcoran, Hudson's Bay Company, 1827-1841*, Martin Comeau examines the uses of the Hudson's Bay Company Archives in paleoclimatology.[4] While these examples use analogue records in their studies, these uses will one day be extended to digital records. This is not only because there is an increase in born digital records, but also because many analogue records are being digitized to facilitate research. The fields of medicine, climatology, and others all use digital records which will need to be preserved. If these digital records are not preserved, studies such as these will not be able to use these records. Digital records are also used by researchers of technology. Users ranging from historians to software engineers are interested in the history of software and other digital technologies.[5] Without digital preservation, researchers will not be able to successfully conduct studies of past and present digital cultures.

---

[2] David A. Snowdon, "Healthy Aging and Dementia: Findings from the Nun Study," *Annals of Internal Medicine* vol. 139, (2003): pp. 450-454.

[3] Natalie Vielfaure, "Medical Records Redefined: The Value of the Archival Record in Medical Research" (Master's thesis, University of Manitoba, 2015).

[4] Martin Comeau, "Archives, Historical Climate Records, and the Climate Observations of Thomas Corcoran, Hudson's Bay Company, 1827-1841" (Master's thesis, University of Manitoba, 2005).

[5] For a brief overview of the historiography of the history of software, please see Martin Campbell-Kelly, "The History of the History of Software," *IEEE Annals of the History of Computing* (2007): pp. 40-51. For a brief overview of the historiography of the World Wide Web, see Megan Sapnar Ankerson, "Writing Web Histories with an Eye on the Analog Past," *New Media and Society* 14 (2012), pp. 384-400.

Digital preservation is not just important to research communities but also various organizations including businesses and governments. Digital records need to be kept by organizations for many reasons. Organizations need to keep records to maintain their corporate memory so that valuable organizational knowledge and expertise for future employees of the company are not lost. Digital records may need to be kept for legal reasons or to defend the organization's reputation by providing its own records as evidence. Digital preservation can also ensure that digital records are properly maintained and organized so that they can be efficiently retrieved by the organization. Finally, digital records may be kept by an organization due to legislative requirements. This is especially true for governments that need to keep records to supply evidence of their activities so that government agencies can be held accountable for their actions.[6]

While digital preservation is important for researchers and organizations, it is also important for individuals. With proper digital preservation, people can preserve their own personal records for as long as they would like. Photographs from a family vacation, video of the graduation of a loved one, blogs and social media: all of these digital records can be preserved by an individual. These digital records can also be passed on through the generations to help preserve a family's memory of past milestones. The reasons that they are kept vary from individual to individual, but being able to preserve them for as long as deemed necessary is a major contribution of digital preservation.[7]

---

[6] Adrian Brown, *Practical Digital Preservation: A How-to Guide for Organizations of Any Size* (London: Facet Publishing, 2013), pp. 19-24.

[7] Several authors have written about personal records and how they are kept/archived. Please see Jordan Bass, "Getting Personal: Confronting the Challenges of Archiving Personal Records in the Digital Age" (Master's thesis, University of Manitoba, 2012); Bass, "A PIM Perspective: Leveraging Personal Information Management Research in the Archiving of Personal Digital Records," *Archivaria* 75 (Spring 2013): pp. 49-76; Sue McKemmish, "Evidence of Me," *The Australian Library Journal* 45:3, pp. 174-187.

Digital preservation is important for the safeguarding of our digital heritage. Academics and researchers need digital records for their studies while businesses and governments need digital records for legal and financial reasons as well as for creating a corporate memory and accountability purposes. Finally, individual people may want to keep certain digital records for a variety of reasons. All of these sectors of our society need to safeguard digital records and as a result, digital preservation is important. It is for these reasons that archivists have developed various digital preservation techniques and systems. Without them, many digital records will either be lost or out of context. It is also important to advocate the importance of digital preservation. In his thesis, *Advocating Electronic Records: Archival and Records Management Promotion of New Approaches to Long-Term Digital Preservation,* Daniel Elves highlights the importance of advocating the need for digital preservation. He reframes digital preservation as not only a technological issue, but one of advocacy and highlighting the necessity of digital preservation.[8] My thesis presents a discussion of the challenges and techniques of digital preservation.

The challenges of digital preservation have been written about in various fields including archival studies, library studies, digital curation, and computer science. There were many projects and studies that were conducted to examine the challenges involved in preserving digital records. In his 1991 article, "Easy to Byte, Harder to Chew: The Second Generation of Electronic Record Archives," Terry Cook examines the archival literature about the hurdles facing digital preservation. Cook divides the archival literature into two generations. The first generation of digital records dealt with the early machine readable records such as punch cards and tapes that were read by mainframe computers. This is different from what Cook terms the

---

[8] Daniel Elves, "Advocating Electronic Records: Archival and Records Management Promotion of New Approaches to Long-Term Digital Preservation," (Master's thesis, University of Manitoba, 2012).

second generation of digital records which are in a wide assortment of file formats and storage media.[9]

Other scholars have identified similar challenges in digital preservation. In his 1988 article, "Computers, Electronic Data, and the Vietnam War," Donald Fisher Harrison discusses the problems with digital records that were created during the Vietnam War by the United States. Harrison explains the difficulty in accessing this data due to some records being dependent on specific software that was used by the American military.[10] During the war, many army units did not keep the records that they had created digitally. This was due to the fact that the National Archives and Record Administration (NARA) did not have a policy regarding the disposition of digital records. Furthermore, many military units and records managers within the Department of Defence saw these digital records as disposable. As a result, many units simply destroyed the records instead of sending them to NARA.[11] The consequences of this was that there was a gap in the historical record of the Vietnam War due to digital records not being preserved.

In his 1993 article, "Electronically Generated Records and Twentieth Century History," historian Ronald Zweig describes many challenges in digital preservation. He identifies the challenges posed by multiple file formats and a dependency of software and hardware as being problematic issues in digital preservation. Zweig also identifies the records management challenge of organizing large volumes of digital records.[12] Margaret Hedstrom identified similar challenges in her 1998 article, "Digital Preservation: A Time Bomb for Digital Libraries." Similar to Zweig, Hedstrom describes the degradation of storage media and other hardware as

[9] Terry Cook, "Easy to Byte, Harder to Chew: The Second Generation of Electronic Records Archives," *Archivaria* 33 (Winter 1991-92): pp. 202-216.
[10] Donald Fisher Harrison, "Computers, Electronic Data, and the Vietnam War," *Archivaria* 26 (Summer 1988): pp. 28-29.
[11] Harrison, "Computers, Electronic Data, and the Vietnam War," p. 26.
[12] Ronald Zweig, "Electronically Generated Records and Twentieth Century History," *Computers and The Humanities* 27 (1993): pp. 74-76.

well as the obsolescence of software and machines as significant challenges to digital preservation. Hedstrom also cited the lack of consistent standards and the large volume of digital records that can be created easily as a major challenge.[13]

Historian Roy Rosenzweig discusses many challenges to digital preservation. In his article, "Scarcity or Abundance? Preserving the Past in a Digital Era," the author mentions several of the challenges described by Hedstrom and Zweig. Rosenzweig discusses how vulnerable aging media is to obsolescence and degradation and the reliance of hardware and software to access digital records.[14] In addition to these technological challenges, Rosenzweig identifies many legal, organizational, social, and economic challenges that he describes as being more challenging than technological issues because they "have disrupted long-evolved systems of trust and authenticity, ownership, and preservation."[15] Because vast quantities of digital records can be easily copied and modified, it becomes challenging to determine which ones are "authentic" and what is unreliable information. The ease of copying and modification also leads to several challenges with copyright because the digital record's rights holder may not allow an archive to make copies of a work for preservation or distribute the record for research purposes.[16] This is in addition to managing the plethora of digital records that are created daily.[17]

There have been many projects that aimed to create solutions for these challenges. A significant early project was the Variables in the Satisfaction of Archival Requirements for Electronic Records Management. This project was conducted at the University of Pittsburgh between 1993 and 1996 by David Bearman and his colleagues. Better known as the Pittsburgh

---

[13] Margaret Hedstrom, "Digital Preservation: A Time Bomb for Digital Libraries," *Computers and Humanities* 31:3 (1997), pp. 189-202.
[14] Roy Rosenzweig, "Scarcity or Abundance? Preserving the Past in a Digital Era," *American Historical Review* 108:3 (June 2003): pp. 740-742.
[15] Rosenzweig, "Scarcity or Abundance," p. 743.
[16] Rosenzweig, "Scarcity or Abundance," pp. 743-746.
[17] Rosenzweig, "Scarcity or Abundance," pp. 757-758.

Project, its main goal was to combine archival and records management theories to ensure that digital records were described with sufficient amounts of metadata at the time of their creation by establishing functional requirements. By doing so, it was hoped that digital records could act as evidence of the activities of their creator.[18] This was done through the creation of functional requirements that were meant to guide in the management of digital records.[19] Combining expertise in multiple fields that included archival, library, and information studies, the project was to accomplish its goals by addressing many issues relating to digital records. These issues related to recordkeeping functional requirements and both the human and technological variables that affect them.[20] In short, the Pittsburgh project wanted to ensure that metadata was added to records upon their creation using functional requirements.

The Pittsburgh Project set out many functional requirements that will allow for metadata to be added to a record when it is created. These requirements ensure that records can adhere to local laws regarding proper recordkeeping. They also require that recordkeeping systems capture records with sufficient metadata and that they are preserved correctly and remain usable for as long as necessary. Finally, this is done to ensure that organizations remain accountable and transparent.[21] While these requirements set out valuable functional requirements that could ensure that records are reliable and authentic, as they stand they are imprecise and do not lend themselves to practical application in the workplace. As a result, the Pittsburgh Project

---

[18] Konrad Krahn, "Looking Under the Hood: Unravelling the Content, Structure, and Context of Functional Requirements for Electronic Record Keeping Systems", (Master's thesis, University of Manitoba, 2012): p. 34. In an ironic twist of fate, most of the original content that was created by the Pittsburgh Project was lost during a server migration at the University of Pittsburgh in the mid-1990s. Much of the original data was recovered through the Internet Archive's Wayback Machine and the University of Pittsburgh's Miranda Nixon in 2008. Please see Krahn, pp. 45-46.

[19] David Bearman, "Record-Keeping Systems," *Archivaria* 36 (1993): pp. 27-32.

[20] Bearman, "Record-Keeping Systems," pp. 27-32; Krahn, "Looking Under the Hood," pp. 34-35.

[21] Since this is not the place for a detailed discussion of the Pittsburgh Project's functional requirements, please see Bearman, "Record-Keeping Systems," pp. 31-32; Krahn pp. 36-43.

developed a series of other publications to supplement these functional requirements. These supplements added more metadata requirements to the Project's findings.[22]

The Pittsburgh Project influenced many follow-up projects in the years following its completion. Among these was the development of the Victorian Electronic Records Projects at the Public Records Office of the State of Victoria in Australia.[23] The Pittsburgh Project reaffirmed the traditional archival principle of provenance by attempting to ensure that digital records have proper metadata. This had the benefit of ensuring that records remained in context and are reliable evidence for the activity of their creators. The work of the Pittsburgh Project also helped facilitate the creation of numerous metadata standards.[24] While the Project was influential, it did have some critics. Terry Cook argued that while the Pittsburgh Project was useful for institutions and other large organizations, it was not as helpful to individuals who did not create records for business reasons.[25] Furthermore, Cook believed that the Project's metadata standards could never provide a substitute for rich archival description.[26] Finally, Cook believed that while Bearman's ideas are helpful, they do not address the realities of working with digital records.[27] Despite these critiques, the Pittsburgh Project was one of the first projects that attempted to find solutions to the challenges of preserving digital records.

Another significant effort to study digital objects was *The Preservation of the Integrity of Electronic Records* project (also known as the UBC Project) that was undertaken at the University of British Columbia between 1994 and 1997. Headed by Luciana Duranti and Terry

---

[22] Krahn, "Looking Under the Hood," pp. 43-45.
[23] David Bearman, "Moments of Risk: Identifying Threats to Electronic Records," *Archivaria* 62 (2006): p. 21; for more information on the Victorian Electronic Records Strategy, please see http://prov.vic.gov.au/government/vers.
[24] Krahn, "Looking Under the Hood," p. 46.
[25] Terry Cook, "The Impact of David Bearman on Modern Archival Thinking: An Essay of Personal Reflection and Critique," *Archives and Museum Informatics* 11 (1997): p. 29.
[26] Cook, "Impact of David Bearman," p. 32.
[27] Cook, "Impact of David Bearman," pp. 31, 33-34.

Eastwood, the UBC Project was funded by a three year grant from the Social Sciences and Humanities Research Council of Canada (SSHRCC). Duranti brought her expertise with diplomatics to the project while Eastwood brought his years of experience as a professional archivist and head of the University of British Columbia's School of Library, Archives and Information Studies from 1981 to 2000.[28] The UBC Project had many goals including the identification of records in digital format and identifying their key parts. The UBC Project also wanted to investigate what kinds of electronic systems produce records and to establish criteria that ensured the authenticity and reliability of records. Finally, the UBC Project wanted to assess the methods of ensuring authenticity and reliability "against different administrative, judicial, cultural, and disciplinary points of view."[29]

The concepts of authenticity and reliability stem from Duranti's research into diplomatics. Diplomatics arose as a method of analysis that determines the authenticity of ancient and medieval documents by examining the record.[30] For Duranti, a record is reliable when it can be shown to be created by an institution or person authorized to do so. A record is authentic when it is what it claims to be and has not been significantly altered.[31] An illustration of reliability and authenticity would be a driver's licence. A driver's licence is authentic when it is what it purports to be and it is reliable when it can be shown to be issued by the proper authority. In the case of my driver's licence, it is authentic because it was issued by the appropriate government agency and it is authentic because it contains information that we can accept as unaltered.

---

[28] Krahn, "Looking Under the Hood," p. 48.
[29] InterPARES, "The Preservation of the Integrity of Electronic Records," http://www.interpares.org/UBCProject/intro.htm.
[30] Luciana Duranti and Heather MacNeil, "The Protection of the Integrity of Electronic Records: An Overview of the UBC-MAS Research Project," *Archivaria* 42 (Fall 2006): p. 47; Krahn, "Looking Under the Hood," pp. 48-49.
[31] Luciana Duranti, "Reliability and Authenticity: The Concepts and their Implications," *Archivaria* 39 (Winter 1994-95), pp. 5-9.

Using these concepts as guidelines, the UBC Project explored many aspects of digital records. It discovered that the reliability and authenticity of records are secured through the establishment of proper record systems and procedures. Authenticity and reliability are also ensured through securing their context and managing them in conjunction with other aspects of the fonds that they belong to. The UBC Project's findings also state that the life-cycle model of records management can be divided into two phases. The first phase is where records are created, used, and stored as semi-active records by an agency and the second phase involves the storage of these records for long term access. Finally, the UBC Project determined that the best agency to manage records is the organization that is responsible for the creation of records.[32]

While these findings were lauded by many, Paul Marsden did have some criticisms of the UBC Project. Marsden believed that while the UBC Project did fantastic work in developing a method of determining reliable and authentic records, its goals could not be realized in many organizations. This was due to the fact that many organizations do not possess the highly centralized administrative structure that the UBC Project recommended for proper management of records.[33] This reality has not improved over a decade later and does not reflect the records management capacities of many record creators.[34] This critique aside, the UBC Project was influential. Many of its findings were applied to the United States Department of Defence to create DoD 5015.2 in 1997. This was the standard for records management applications and how they could satisfy the legal and foundational requirements of bodies within the Department of Defence.[35] Also, the UBC Project's Luciana Duranti created the International Research on

[32] InterPARES, "The Preservation of the Integrity of Electronic Records," http://www.interpares.org/UBCProject/intro.htm.
[33] Paul Marsden, "When is the Future? Comparative Notes on the Electronic Record-Keeping Projects of the University of Pittsburgh and the University of British Columbia," *Archivaria* 43 (Spring 1997), p. 171.
[34] Krahn, "Looking Under the Hood," p. 61.
[35] Krahn, "Looking Under the Hood," p. 67.

Permanent Authentic Records in Electronic Systems (InterPARES), which is focused on the authenticity and preservation of digital records.[36]

The Pittsburgh and the UBC projects were early efforts to understand digital records. While they share many similarities, there were some key differences between the two projects. The UBC Project was heavily grounded in traditional archival theory where the record creators would send their records to an archive which would then arrange and describe the records according to the principle of provenance. In other words, the archives is a passive recipient of the records that were made by a creator. The Pittsburgh Project took a slightly different approach. For them, the archives should take an active role to ensure that metadata was added to records upon their creation. By doing so, the archives would be able to ensure that records were properly described and in context.[37] Despite such differences, these projects remain influential in digital preservation.

Furthermore, these projects have contributed to the development of standards and systems that have become widely used by archivists. It is also important to note that other groups besides archivists have been working through these challenges. As we will see with the Open Archival Information System in Chapter Two, this standard was not developed by archivists but by the Consultative Committee for Space Data Systems. Work by archivists and others groups have also influenced the development of digital preservation systems such as Archivematica and metadata standards such as PREMIS.

Despite this work, no one has yet found a permanent solution to the challenges of digital preservation nor are they likely to. There are many reasons why these challenges remain a persistent part of our digital heritage and many of them have been raised by the authors above.

[36] Krahn, "Looking Under the Hood," pp. 61-62.
[37] Krahn, "Looking Under the Hood," pp.52-53.

There is a wide variety of digital records being created every day and they are in an assortment of file formats and storage media. These records are also reliant on software and hardware to be understood by people. This makes selecting digital records and preserving them a challenge for archivists that is compounded by the fact that archives rarely have the additional funding to hire additional staff to address these digital preservation concerns. As a result, archives have a difficult time adapting to the preservation challenges that are presented with rapidly evolving technologies. This lack of support has created an environment where solutions are difficult to develop.

As we will see later in this thesis, digital records require archivists to be proactive by managing file formats, selecting significant properties, and managing metadata. This work requires the archive to be an active participant in a record's life cycle and involves forming partnerships with various departments including IT and records management. For records that have long retention periods, preservation actions may have to be taken before ingest into the archives happens. Archives should not be passive recipients of records as per the traditional "custodial" model of archives. Rather, a "pre-custodial"[38] approach is required to ensure that the needs of the archives are being met. This increased involvement of the archives will have the benefit of making it more prominent in an organization, which in turn may lead to increased funding. With more funding, the archives can afford to hire more staff who can devote more time to addressing digital preservation challenges. As we will see, these challenges are complex and often require additional staff resources. What follows is a description of why these

[38] I am using Adrian Cunningham's definition of pre-custodial which is defined as having the archives take a more active role in records management. This is in contrast with the "custodial" view where archives are passive recipients of records. For more information, see Adrian Cunningham, "Waiting for the Ghost Train: Stratagies for Managing Electronic Personal Records Before its too Late," *Archival Issues* 24:1 (1999): pp. 55-64; Cunningham, "The Archival Management of Personal Records in Electronic Form: Some Suggestions," *Archives and Manuscripts*, 22:94 (1994): pp. 94-105.

challenges are still present and what has been done to address them. I will also explain the reasons why permanent solutions to these challenges are difficult to develop.

Chapter One

The Challenges of Digital Preservation

There are many challenges in digital preservation that hamper the efforts of information professionals to preserve digital records. These challenges affect various aspects of an organization. Digital preservation challenges can ruin an organization's corporate memory. This memory becomes vital when the organization deals with highly specialized knowledge that needs to be maintained over long periods of time. This data needs to be reused by staff and others for the company to handle various situations. These include major data loss, legal cases, and making information available in accordance with laws. A commonly cited statistic is that 90 percent of businesses that suffer a major loss of data go out of business within two years.[1] Clearly, these preservation challenges can cause significant damage to organizations.

These challenges stem from the nature of the digital records. Unlike analog objects, digital objects are composed of a series of ones and zeroes called a bitstream in addition to being stored on a physical medium such as a DVD or hard drive. This physical medium requires specific hardware to access its stored files, which in turn requires the correct software that can read the bitstream and render it into a format that is readable to humans.[2] In many ways, this process of rendering digital records is analogous to certain machine readable analog records. An example is a film reel and a projector. By itself, it is difficult to understand the contents of a film reel if it is not projected onto a screen for viewing. It is only through the aid of a projector that the contents of a reel become clear as they are projected onto a screen. A digital object is akin to the content of a film reel. Alone, it is unreadable, but with the aid of hardware and software (i.e.

---

[1] Adrian Brown, *Practical Digital Preservation: A How-to Guide for Organizations of Any Size* (London, 2013): pp. 20-22.
[2] Mike Kastellec, "Practical Limits to the Scope of Digital Preservation," *Information Technology and Libraries* 31:2 (2012): p. 64.

a computer system) the digital object is rendered into something that can be understood in the same way that a film is when it is projected onto a screen.

This reliance on hardware and software leads to numerous challenges in digital preservation. Storage media can fail and degrade the digital records written to it. The hardware required to read the storage device can become damaged or obsolete as time goes by and software can become corrupted or be rendered obsolete by newer programs. The fragility and speed of technological advances create numerous challenges for information professionals who are trying to preserve digital records.

While digital preservation has many challenges that are technological in nature, it is necessary to mention that digital preservation is just as much a human problem as it is a technological one. Through their use of digital technology, people create large quantities of digital records. These records are stored in various systems of organization. These systems of organization can range from highly organized file structures to ad hoc and random methods of organization.[3] This requires that archivists appraise copious amounts of digital records that are stored in diverse storage systems. Once a group of records has been selected for preservation, archivists have to determine the significant properties of the records that must be preserved and ensure that they are described properly so that the records will remain in context and thus understandable. Finally, there is a whole suite of legal and financial constraints on digital preservation that make it challenging. In short, digital preservation is just as much of a human problem as it is technological.

In this chapter, I will be examining these challenges and explain why after many years they still remain. I will begin by historicizing these challenges to show that they have been

---

[3] For more details on how people organize their personal records, please see Bass, "A PIM Perspective."

around for as long as digital records have been in existence. I will then examine the challenges of digital preservation individually to explain how they will remain challenges to the preservation of digital records.

Challenges of Digital Preservation

The many challenges to digital preservation can be divided into two groups: technological and human. Technical challenges stem from the reliance on the hardware and software that is required to store and read digital records. Problems are caused by the degradation of the bitstream and storage devices as well as the obsolescence of hardware and software. Despite this reliance on technology, we should not forget the effect of people in digital preservation. It is for humans, not machines, that we endeavor to maintain the original meaning and context of a digital record. It is also through human actions that we attach value to digital records and the properties that they contain. All of this leads to two questions: what is considered valuable enough to preserve and what properties are considered significant to a digital record? The answer to these questions lies within our own conceptions of value. Perceptions of value provide a significant challenge to digital preservation efforts because of their subjective nature. This can result in loss of what some individuals may consider to be valuable digital records or properties of a digital record. In short, human challenges of digital preservation consist of selecting those records valuable enough to preserve, identifying their significant properties, and knowing how to preserve them.

Technological Challenges

There are many technological challenges to digital preservation. These include deterioration in the storage media and the hardware required to read these devices. Furthermore, these devices are vulnerable to becoming obsolete. Software is also exposed to these same challenges. Finally, the digital objects themselves are susceptible to deterioration. What follows is an examination of these challenges and why they remain after several decades of digital preservation.

Digital objects are composed of a string of zeroes and ones called a bitstream, which is stored on physical media such as a flash drive.[4] This string of binary code is unintelligible to humans and is only readable with the intervention of hardware and software (which is itself a digital record as it too requires the correct combination of hardware and software to run). While both of these will eventually become obsolete, the bitstream of digital objects degrades if it is not properly stored. This is a form of degradation that is referred to as bit rot.[5] Indeed, the degradation of both the digital record and the storage media on which they are stored is a significant part of the technological challenge of digital preservation.

There are various causes of degradation including viruses, natural disaster, and hardware deterioration. In addition to degradation, there is also destruction of the bitstream through human actions. These range from accidental deletion to destruction through war, vandalism, purposefully removing evidence, and other harmful acts.[6] Furthermore, hardware manufacturers often exaggerate the lifespan of their devices when in reality, they have no way of ensuring that the data stored on the device will last that long.[7] Despite the lack of a perfect storage solution,

---

[4] Kastellec, “Practical Limits,” pp. 63-64.
[5] Vinton G. Cerf, "Avoiding ‘Bit Rot’: Long-Term Preservation of Digital Information," *Proceedings of the IEEE* 99:6 (June 2011): pp. 915-916.
[6] Kastellec, “Practical Limits,” p. 64.
[7] David S. H. Rosenthal "Keeping Bits Safe: How Hard Can it Be?" *Communications of the ACM* 53:11 (November 2010): pp. 48-51.

Rosenthal believes that our current methods of storage are the most feasible approach archivists have as it is an impossible task find the perfect storage device.[8]

Even though it is clear that hardware cannot keep a digital record in perpetuity, measures have been put in place to ensure that data can be preserved for as long as necessary. A common strategy to combat hardware errors is called data redundancy. This involves making multiple copies of digital records and regularly auditing them to ensure that they have not been corrupted. If damaged files are found, they would be replaced with a clean copy of the digital object.[9] While this method does work, it is vulnerable to many factors. First, it is expensive in terms of financial and human resources to maintain servers and other storage devices over long periods of time. As the amount of data an archive stores increases, so does the amount of storage required and therefore the cost. This results in even more storage devices being needed. People also need to audit these records to ensure that there is no corruption in the copies and as the archive grows, so does the amount of work involved. The resulting time between audits increases, which naturally leads to a higher chance of data corruption.[10] All of these challenges are in addition to the hardware failures, which are inevitable.

Regardless of these difficulties, it is absolutely imperative that copies of digital records are made to safeguard against the eventual degradation of hardware and digital objects. While Rosenthal is correct that data redundancy cannot solve the problem, it is also a necessary response to it. Multiple copies of records need to be kept to safeguard against the eventual degradation of storage devices. Archivists have no way of ensuring the reliability of storage media since there is no clear experiment that can test how long data can be stored on that device.

[8] David S. H. Rosenthal, "Bit Preservation: A Solved Problem?" *The International Journal of Digital Curation* 5:1 (2010): p. 140.
[9] Rosenthal, "Bit Preservation," p. 136; Rosenthal, "Keeping Bits Safe," p. 51; Kastellec, “Practical Limits,” p. 64.
[10] Rosenthal, "Bit Preservation," p. 136; Kastellec, “Practical Limits,” pp. 64-65.

If one was developed, by the time that the experiment was completed (it would more than likely take many years of monitoring), the results would be no longer useful because a new device would become available and the old ones would become obsolete as new storage devices are constantly appearing on the market.

It is for these reasons that the problem of degradation is still prevalent today. All physical objects eventually degrade with age and computer hardware is no exception. As computers and storage devices degrade, so does that data that is contained within them. When the bitstream is corrupted, it will become difficult for hardware and software to interpret it and render it in a format that is intelligible to the human eye. While it is impossible to determine when a natural disaster will occur, backing up and auditing data is an excellent way to ensure that the bitstream of a digital record is able to survive a hardware failure. Suppose, however, that the storage device that a digital record is written on is maintained with no damage to the bitstream. How will this data be accessed? Technological advancement is increasing at an exponential rate and older technologies are becoming obsolete. Floppy discs are no longer used and CDs, DVDs, and other media are becoming less common.

Obsolescence is one of the major obstacles in the technological challenges of digital preservation. Obsolescence occurs when the hardware and software required to interpret the bitstream is no longer available due to the rapid pace of technological change.[11] Unlike paper records, digital records require physical hardware and digital software to be understood by humans. Obsolescence adds an extra layer of challenge to the difficulty of preserving the bits on a storage device. A bitstream can be perfectly preserved on a storage device if it is properly

[11] Brown, *Practical Digital Preservation*, pp. 202-203; Kastellec, "Practical Limits," p. 64.

maintained and backups are made in case of an emergency. If this bitstream is stored on a medium that can no longer be read, it will be difficult to access the data.

An example of the difficulties obsolescence can impose on an organization is found in Charles Levi's article "Five Hundred 5.25-inch Discs and One (finicky) Machine: A Report on a Legacy e-records Pilot Project at the Archives of Ontario." In this article, Levi explains the difficulties that the Archives of Ontario encountered when it processed numerous 3.5 and 5.25 floppy disks from the 1980s and 1990s that were filed with paper records. The difficulty was that the archives did not have a computer that could read a floppy disk because the technology was obsolete. Furthermore, when the Archives of Ontario did eventually get access to the data on the floppy disks, it had difficulty converting some of the files into newer formats because they were created in a format that was not widely supported anymore. This was mainly due to the fact that there were few standardized formats in the 1980s and 1990s.[12] There are many similar experiences in both business and private life. Most people can recall an instance when they found an old floppy disk and have no way of accessing its files or they found an old computer file that cannot be opened because it is in a format that is no longer supported.

As can be seen in the examples above, obsolescence is not just a hardware challenge but a software one because file formats and software become obsolete due to the rapid changes in software technology. There are various reasons as to why software and file formats can become obsolete. Much like hardware, file formats can become obsolete as newer versions of software become unable to read older formats. David S.H. Rosenthal argues that file formats become obsolete when they fail to become dominant while a market is in its early stages. In an early market, there are often various file formats competing for the dominant share of the market.

[12] Charles Levi, "Five Hundred 5.25-inch Discs and One (finicky) Machine: A Report on a Legacy e-records Pilot Project at the Archives of Ontario," *Archivaria* 72 (2011): pp. 239-246.

Eventually, one or a select few formats become widely used by most of the community and achieve market dominance at the expense of other competing formats. Once the market matures with only a select few formats becoming dominant, all others quickly become obsolete.[13] Rosenthal further argues that many of these unsuccessful formats are often not worth preserving since they were not widely used by a particular community.[14]

While formats that fail to capture a significant portion of a market have a higher tendency to become obsolete, these formats are still worth preserving. Indeed, Michael Mahoney argues that there are too many histories that focus on the success stories of specific pieces of software at the expense of the more numerous software systems that failed to become the dominant product.[15] Different communities create a piece of technology for a variety of reasons. Technology does not have a mind of its own but was created by people to solve a specific problem. Mahoney argues that the history of computing too often only focuses on the success stories of the computer and treats success as revolutionary. The contributions of the products and ideas that existed before the computer and the human agency involved are given less attention. Instead of focusing on the hype that the computer had generated, Mahoney argues that we should focus on why the machine was created and how it was used. By doing so, we will be able to keep the machine in context.[16]

Context is incredibly important in digital preservation. Knowing why something was developed in a specific way is integral to understanding the nature of the record. In particular, it is important to understand the technological landscape in which a particular program was

---

[13] David S.H. Rosenthal, "Format Obsolescence: Assessing the Threat and the Defenses," *Library Hi Tech* 28:2 (2010): pp. 196-200.
[14] Rosenthal, "Obsolescence," p. 197.
[15] Michael Mahoney, "What Makes the History of Software Hard," *IEEE Annals of the History of Computing* (July-September 2008): p. 12.
[16] Michael Mahoney, "The Histories of Computing(s)," *Interdisciplinary Science Reviews* 30:2 (2005): pp. 119-127.

developed. This includes understanding the technologies that existed before and competed with a particular program. To return to Rosenthal's remark that many of these early file formats and programs were not valuable because they failed to be widely adopted, it is necessary to understand and to preserve these digital records because that will help us understand the digital landscape that led to the development of other programs and also the digital environments in which they were deployed.

With all of this being said, how do we mitigate the challenges posed by obsolescence? The digital preservation community has developed many methods that reduce the threat posed by obsolescence. Digital records and storage media can be migrated to newer versions to postpone obsolescence. Alternatively, many copies of digital records will be made and checked for errors to ensure that there is no data loss. This process is called data redundancy. Archives can also migrate the formats of digital records to standardized formats that are widely used. Finally, digital records can be rendered in an operating environment similar to the one that it was originally designed for in a process called emulation.[17] It is crucial to note that these measures do not end the difficulties posed by obsolescence. Rather, they postpone the eventual obsolescence of a digital record to ensure that it remains accessible. Technology will continue to advance and cause whatever format or storage medium that the digital records were transferred into to eventually become obsolete. The methods described here are temporary measures that can assist archivists with maintaining accessibility of digital records.

There are many forms of migration but they all address the challenges associated with obsolescence and degradation of hardware, software, and file formats by allowing digital objects to be readable. This is done by changing either an obsolete storage medium or file format to

[17] Kastellec, "Practical Limits," p. 64.

something that is current and readable on current hardware.[18] This can involve copying a digital object's bit stream from an outdated storage medium to one that is currently used. This process is called refreshment.[19] For migrating file formats, the outdated format is changed to one that is current.[20] One example of this is simple version migration. This involves migrating an obsolete file format to updated versions that are within the family of products or formats.[21] An example of this is changing from an old version of Microsoft Word to a newer one.

There are some problems with migration. Jeff Rothenberg believes migration is simply "wishful thinking" on the part of archivists as each migration is costly in terms of money and staff resources and also runs the risk of data loss.[22] Migration is not a permanent solution to the technological challenges of digital preservation but a temporary solution. Migration would have to be performed many times in the lifespan of a digital record. With each migration, careful analysis of the formats needs to be conducted to decrease the risk of data loss. This makes the process of migration extremely time consuming and labour intensive.[23] Furthermore, this labour intensive process needs to be conducted every time a migration occurs.[24] This care must be taken to mitigate the risk of losing valuable data during the transfer. When older formats are migrated to current ones, some functionalities such as fonts and formatting of the originals can

---

[18] Ross Harvey, *Preserving Digital Materials*, (Berlin: De Gruyter, 2012): p. 156; Kastellec, “Practical Limits,” p. 64.
[19] Harvey, *Preserving Digital Materials*, pp, 142-143.
[20] In some circles, this form of migration is often called transformation since the main goal is to change the bitstream of the digital object. Please see David Giaretta, *Advanced Digital Preservation*, (Springer: Berlin, 2011): pp. 200-202.
[21] Kenneth Thibodeau, "Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years," in *The State of Digital Preservation: An International Perspective* (July 2002): p. 23.
[22] Jeff Rothenberg, "Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation, A Report to the Council on Library and Information Resources," (January 1999): pp. 13-16.
[23] United Nations Educational, Scientific and Cultural Organization, "Guidelines for the Preservation of Digital Heritage," (March 2003): p. 135.
[24] Harvey, *Preserving Digital Materials*, p.159.

become corrupted and unable to be rendered properly.[25] Therefore, care must be taken when migration is performed to ensure that there is little chance for data loss to occur.

Migration is often used with another digital preservation strategy called normalization.[26] Normalization involves converting files into preservation formats as soon as they are acquired by the archive.[27] The purpose of this strategy is to facilitate preservation in the future by having all files acquired by the archive be in a format that is stable and less likely to become obsolete in the immediate future.[28] A good example of normalization can be found in Kye O'Donnell's article "Taming Digital Records with the Warrior Princess: Developing a Xena Preservation Interface with TRIM." In this article, O'Donnell describes an implementation of Xena (which stand for XML Electronic Normalizing for Archives) and its pros and cons. Xena is an open source normalization tool that was developed by the National Archives of Australia. Xena identifies a file and then generates metadata while it is converted into an open format. Xena also encases the digital object in XML to produce a XENA file.[29] As with migration, however, normalization is a labour intensive process and costs money to convert files into standardized preservation formats.[30] There is also the risk of losing functionalities and the original look and feel of the records when it is normalized.

Another safeguard against obsolescence is format management. This involves developing a plan that limits the variety of file formats that an archive will accept to facilitate

---

[25] Brown, *Practical Digital Preservation*, pp. 209-210; for an example of how functionalities can be lost through migration, see Jeff Bennion, "The Rage-inducing Word versus WordPerfect Debate," http://abovethelaw.com/2014/06/the-rage-inducing-word-versus-wordperfect-debate/, last accessed September 23, 2016.

[26] Harvey, *Preserving Digital Materials,* p. 158.

[27] Brown, *Practical Digital Preservation*, p. 211; Harvey, *Preserving Digital Materials*, p. 158; Kastellec, "Practical Limits," p. 64.

[28] Brown, *Practical Digital Preservation*, p. 211; Harvey, *Preserving Digital Materials*, p. 158.

[29] Kye O'Donnell, "Taming Digital Records with the Warrior Princess: Developing a Xena Preservation Interface with TRIM," *Archives and Manuscripts*, 38:2 (November 2010): pp. 37-60.

[30] Kastellec, "Practical Limits," p. 64.

preservation. Many file formats require specific hardware and software to be rendered. When these become unavailable through obsolescence or other challenges, the risk that the digital object would be lost increases. Furthermore, many formats are proprietary and that will run the risk of obsolescence if the company that supports them goes out of business. Through format management, an archive can have greater control over the file formats in their archive by limiting the formats that will be accepted to those that are able to be effectively preserved by the archive.[31] To summarize, format management involves limiting the number of file formats that will be accepted into an archive to those that can be effectively preserved.

There are many examples of format management. One such example is at Library and Archives Canada (LAC) when it created the Local Digital Format Registry (LDFR). The LDFR outlines what file formats and migration pathways for content that LAC will accept by analysing whether a format can be preserved for an extended period of time. LAC also identified the need for the list of file formats to change over a period of time as file formats become obsolete.[32] The Library of Congress has also developed its own format registry that outlines what formats it will accept into its archives.[33] Format management is an important part of digital preservation as it assists in managing the large volume of digital records that are constantly being created. By maintaining formats that are widely accessible, less likely to become obsolete in the immediate future, and ideally are free from the restrictions of proprietary software, format management facilitates preservation. In short, carefully managing file formats aids digital preservation by reducing the stress caused by preserving multiple file formats.

---

[31] Harvey, *Preserving Digital Materials*, pp. 143-154.
[32] Library and Archives Canada, *Local Digital Format Registry*, p. 1.
[33] Library of Congress, *Recommended Format Specifications*, http://www.loc.gov/preservation/resources/rfs/TOC.html, last accessed September 23, 2016.

Another method of digital preservation is emulation. Emulation involves using a program called an emulator to recreate the computing environment that a digital record would be able to run.[34] Emulation is distinct from preservation methods such as migration and normalization in that it does not change the bitstream of the digital object. An example of this is a Super Nintendo emulator on a computer that can emulate the operating environment of a Super Nintendo to run games from that system. The video game community has made extensive use of emulation to run video games from older consoles. In their article, ""Keeping the Game Alive: Evaluating Strategies for the Preservation of Console Video Games," Mark Guttenbrunner, Christoph Becker, and Andreas Raubner suggest that emulation is needed for the preservation of video games because, unlike video or text files, video games are meant to be interactive. They are not meant to be watched or read butto be played. When a preservation method such as migration or normalization is used, they change the bitstream of the numerous digital objects that make up a video game. As a result, the original game is changed from what its creator had intended it to be and can become unplayable. Also, Guttenbrunner et al. found that there is an added challenge of preserving video games that require specific attachments, such as the bongo drums that are required to play *Donkey Konga*.[35]

It is important to keep in mind, however, that emulation is far from perfect. Mahoney believes that emulators have several deficiencies. While emulators can "convey some sense of programming by executing the code, what appears on the screen is merely a simulation of the physical device, omitting entirely the spatial, temporal, and social experience of writing programs and having them punched up and run, or of interacting with the system as a user."[36]

---

[34] Kastellec, "Practical Limits," p. 64.
[35] Mark Guttenbrunner, Christoph Becker and Andreas Raubner, "Keeping the game alive: evaluating strategies for the preservation of console video games," *International Journal of Digital Curation* 5:1 (2010): pp. 64-90.
[36] Mahoney, "Software," pp. 14-15.

Through my personal experience using various types of video game emulators, I can confirm that this is indeed the case. The sound and graphics may be different from the original game not to mention the way in which someone would play it. For example, someone might play a Super Nintendo game on a PC using a Playstation controller because it is one of the few that can easily connect to a computer. Depending on the emulation program, the sound and graphics would vary in quality. I can say that video game emulators have improved significantly over the years, but they are far from perfect.

Despite these critiques, emulation sometimes allows archivists to preserve the look and feel of a digital object when it is considered to be an important aspect of the digital record. In their article, "A Comprehensive Approach to Born-Digital Archives," Laura Carroll, Erika Farr, Peter Hornsby, and Ben Ranker provide another example of the uses of emulation. They describe the use of emulation for the Salmon Rushdie Collection at the Emory University Libraries. The Rushdie Collection contains records that are both analogue and born digital. The staff at the Emory Libraries wanted researchers to be able to view these digital records in the same operating environment that Rushdie would have used himself. This would allow researchers to view the computing environment that Rushdie wrote in and experience its differences compared to current operating systems. To accomplish this, the staff at Emory created an emulator that can replicate the operating environment of Rushdie's old computer. As a result, researchers can view these digital records in the same way that Rushdie would have viewed them. Emory also had to ensure that many of the records that Rushdie did not want accessible to the public were either redacted or removed from the emulation entirely.[37] In short,

[37] Laura Carroll, Erika Farr, Peter Hornsby, and Ben Ranker, "A Comprehensive Approach to Born-Digital Archives," *Archivaria* 72 (Fall 2011): pp. 61-92.

emulation provides a way for researchers to interact with the digital records in the way that they originally appeared when they were first developed.

While emulation provides a means to simulate the operating environment of a digital record, why should we not just maintain the hardware and software indefinitely to maintain working order? The advantages preserving the original systems are that we would not have to worry about obsolescence because we are preserving the hardware and software required to render the digital record. The original experience of using the machine would also be preserved. The challenge with this method, however, is that hardware degrades. Plastics break, hard drives fail, machine parts burn out, the list can go on. Eventually new parts would have to be used to replace failed ones and this can lead to a loss of data. Assuming no technological failures happen, there is the risk of losing the technical knowledge of using the machine. In her article, “Personal computers, microhistory, and shared authority: Documenting the inventor-early adopter dialectic,” archival scholar Patricia Galloway describes the challenges of using her old Kapyro II computer after several years of sitting in her basement. After fixing the hardware issues and making the computer operational, she found that she had lost the knowledge of using the command line interface.[38] For Galloway, this lack of knowledge is a significant preservation challenge with the original operating environment as the skills required to use these systems with any sort of efficiency are often lost. Therefore, documentation about the use of the system needs to be preserved as well.[39]

This last point is a significant challenge for both systems preservation and emulation. If the knowledge to use these old systems is lost, the digital records that rely on this system would

[38] Patricia Galloway, "Personal computers, microhistory, and shared authority: Documenting the inventor-early adopter dialectic," *IEEE Annals of the History of Computing* 33:2 (2011): pp. 60-61.
[39] Galloway, “Personal Computers,” pp. 64-65.

be difficult to access and/or use. An example of this would be using a pre-GUI operating system with little to no knowledge of command line interfaces. If a researcher does not have the required knowledge to operate such a system, digital records would be inaccessible. Another point to consider is whether the majority of researchers would want to learn how to use these old systems to access a handful of digital records. Unless researchers specifically want to view the operating environment, it would encumber their research if they had to navigate obsolete computer technologies to find one digital record. Documentation will aide in learning old computing environments as it would provide much needed instructions on machine operation but, as Galloway mentions, much of the documentation has been lost.[40] Issues such as these make systems preservation and emulation challenging to accomplish.

The methods and plans outlined above provide some means to mitigate the challenges of digital preservation but they should never be used alone. Instead, specific methods should be used to preserve specific digital records. None of the methods described above can permanently solve any of the challenges of digital preservation. When they are used properly, however, they can allow for digital objects to remain accessible for as long as necessary. Every single one of these methods works for specific situations and often require the use of other methods to be successful. An example is using migration, format management, and data redundancy to preserve a group of digital records. When a digital record is acquired in one of the acceptable preservation formats (format management), an extra copy is made to safeguard against hardware failure (data redundancy). When the format is about to become obsolete, a copy is migrated to a newer format so that people can access it (format migration). When the storage medium becomes obsolete, the digital objects are migrated to a new storage device that can be read by

[40] Galloway, “Personal Computers,” pp. 64-65.

current machines (media migration). A successful digital preservation program needs to use many of these methods in tandem to properly preserve these digital records. Each of these methods has specific uses and knowing when to use each one is an important aspect of digital preservation that will be discussed in the next section.

Human Challenges

While related to technological challenges, human challenges do not necessarily deal with issues of technology. Instead, these challenges stem from such difficulties as determining which digital records out of the multitude that are created daily are worthy of preservation. As seen with the technological challenges, many of the preservation methods that are currently being used involve changing the bitstream of a digital object. Therefore, it becomes necessary to determine what aspects of a digital record must be maintained. This requires archivists to exercise their judgement in selecting what aspects of the record are worth preserving. While everyone can agree that the preservation of contextual information about the creation of the record is of significant importance, there will always be different answers to the question of which digital records should be preserved and what aspects of these ought to be kept. In short, our subjective notions of value affect what we consider to be valuable digital records.

Digital preservation faces many legal hurdles in regard to copyright. As we have seen above, many digital preservation strategies involve making changes to and/or copies of digital records that may be protected under copyright. Copyright law existed well before the advent of digital records and current digital technologies. These laws were originally intended for analogue records,[41] which are preserved by ensuring that they change as little as possible

[41] Kastellec, "Practical Limits," p. 66.

throughout their lifespan.[42] Simply purchasing the digital record to preserve it is often not enough to avoid possible copyright violations as this only grants the end user permission to access the digital content in question.[43] To add further complexity, copyright laws vary from nation to nation.[44] Generally speaking, archives can only make additional copies of a digital record if it is not protected by copyright or if the archives owns the copyright or has permission from the copyright owner.[45]

There are many reasons why copyright restrictions are problematic for digital preservation. As described earlier in this chapter, digital records are less permanent than paper records because they rely on hardware and software to make them readable to users. This challenge results in archivists needing to perform preservation actions much sooner than they would for paper records. If an archivist could not get approval from the rights holder, digital preservation actions such as migration and data redundancy cannot be performed as this would violate copyright laws by changing and making additional copies of an item without the copyright holder's permission. Furthermore, copyright issues do not only involve the content of digital records but also the software required to access them. An example of this is modifying an operating system so that it could run on current hardware. If the archives does not have permission from the operating system's copyright holder, the archives would be breaking the law

---

[42] Ross Harvey and Martha Mahard, "Mapping the Preservation Landscape for the Twenty-First Century," *Preservation, Digital Technology & Culture*, 42:1 (2013): p. 9.
[43] Kastellec, "Practical Limits," pp. 66-67.
[44] An article that highlights the differences in copyright legislation regarding digital content is David Anderson, "Preserving Europe's Digital Cultural Heritage: A Legal Perspective," *New Review of Information Networking* 18:1 (May 2013): 16–39. While this article only focuses on European legislation, it provides valuable insight into the differences in how copyright is handled in various countries and how that relates to digital preservation. Another source that provides information on copyright and digital locks from a Canadian perspective is Michael Geist in his blog: http://www.michaelgeist.ca/tag/digital-locks/.
[45] June M. Besk, Jessica Coates, Brian Fitzgerald, Wilma Mossink, William G. LeFurgy, Adrienne Muir, Mary Rasenberger, Christopher D. Weston, "Digital Preservation and Copyright: An International Study," *The International Journal of Digital Curation* 3:2 (2008): p. 105.

by modifying the software. Legal questions such as these are complex as they require a solid understanding of both digital preservation practices and nation-specific copyright legislation. It is beyond the scope of this thesis to determine the "correct" way to handle copyright challenges as there is no single way of coping with them. My goal here is to raise these issues for future consideration.

When a digital record has been selected for preservation, metadata is used to keep it in context. Metadata is the information that describes aspects of a digital record including information on the record's creator, date of creation, preservation information, and much more.[46] The archival community has often relied upon description to ensure that records remain intelligible for future users by recording the record's provenance or context in which it was created. For digital records, metadata assists in the preservation of a record's context by ensuring that its authenticity and provenance are maintained.[47] This is done through the use of various kinds of metadata types including administrative (data on the rights, preservation, and ownership of a record), structural (data on the internal structure of a record and how it relates to similar content), technical (data on the technical specifications of the record), and descriptive (data on the history of the record).[48]

Despite its importance, the need to collect metadata is often ignored, which in turn leads to the loss of valuable data.[49] A lack of consensus is also apparent in the preservation community as there is a wide array of metadata standards that differ in terminology and

---

[46] P. Ramesh, J. Vivekavardhan, and K. Bharathi, "Metadata Diversity, Interoperability and Resource Discovery Issues and Challenges," *DESIDOC Journal of Library & Information Technology* 35:3 (April 22, 2015): p. 193.
[47] Andrew Wilson, "How Much is Enough: Metadata for Preserving Digital Data," *Journal of Library Metadata* 10 (2010): pp. 207-210. It should be noted that metadata is not a substitute for archival description. Please see Cook, "Impact of David Bearman," p. 32 on his critiques of the Pittsburgh Project.
[48] Ramesh et al., pp. 195-196.
[49] Amelia Breytenbach and Ria Groenewald, "The Use of Metadata and Preservation Methods for Continuous Access to Digital Data," *The Electronic Library* 29:2 (2011): p. 237.

structure.[50] Currently, there is no single metadata standard that adequately describes all aspects of a digital record. This creates significant problems when preservation professionals assign metadata to a record as many standards need to be used together. This often leads to inconsistencies in terms that need to be harmonized by archivists.[51] In short, metadata is an integral part of ensuring that digital records are adequately described but the disparate metadata standards and lack of consensus make describing records in a consistent format difficult.

Another area of digital preservation where there is a lack of consensus among information professionals is selecting which digital records are worthy of preservation. Each day, many different kinds of digital records are created. Each one is created to fulfill a variety of functions, including computer games, movies, websites, pictures, emails, music, spreadsheets, databases, word processed documents and many more. While each of these digital records is useful in its own way, it is impossible to preserve everything. The amount of digital content in existence is staggering and too much for any one institution to manage. How, then, does one determine which digital records will be saved and what will be allowed to fade into obscurity?

Many scholars have stressed the importance of appraising large amount of digital records to determine what is useful for a given designated community. Both Harvey and Kastellec agree that managing the large volumes of digital records is an important challenge for digital preservation.[52] Kastellec outlines several models for selecting digital records:

> Models of selection for digital objects can be plotted on a scale according to the degree of human mediation they entail. At one end, the *selective* model is closest to selection in the analog world, with librarians individually identifying digital objects worthy of digital preservation. At the other end of the scale, the *whole domain* model involves minimal human-mediation, with automated harvesting of digital objects. The *collaborative* model, in which archival institutions negotiate agreements with publishers to deposit content, falls somewhere between these two extremes, as does the *thematic* model, which can

[50] Ramesh et al., "Metadata Diversity," pp. 196-198.
[51] Ramesh et al., "Metadata Diversity," p. 198; Wilson, "How Much is Enough," pp. 212-214.
[52] Harvey, *Preserving Digital Materials*, p. 2; Kastellec, "Practical Limits," p. 67.

> apply either selective- or whole-domain-type approaches to relatively narrow sets of digital objects defined by event, topic, or community.[53]

While each of these approaches has its benefits, they also have disadvantages. By keeping everything, information professionals would be unable to effectively manage the staggering amounts of digital records in their care. Likewise, carefully selecting which records to acquire and which to refuse often leads to an appraisal backlog as it involves large amounts of staff resources to parse through terabytes of data.[54] Many institutions, however, do not have the human resources to be able to sort through large amounts of digital records. In their article, "Overwhelmed to Action: Digital Preservation Challenges at the Under-Resourced Institution," Amanda Kay Rinehart, Patrice-Andre Prud'homme, and Andrew Reid Huot outline many of the challenges that digital preservation presents. In addition to lack of funding, staff often face a lack of training, change fatigue,[55] and a lack of engagement from major decision makers.[56]

Selecting digital records is a form of archival appraisal. Archival appraisal is the act of selecting the records that will be stored in an archive and by extension, it also involves selecting which materials will not be archived, usually destroyed, and thus lost to society. Much has been written on appraisal and the many theories and methods of conducting it.[57] Appraisal theories such as Macroappraisal, Documentation Strategy, and others have been influential. Due to the large quantities of digital records that are created, appraising digital records has provided many

---

[53] Kastellec, "Practical Limits," p. 67.
[54] Kastellec, "Practical Limits," p. 67.
[55] The authors defined change fatigue as the stress that is caused from changes to both the office environment (i.e. the coming and going of staff, building maintenance, moves, etc.) and technology.
[56] Amanda Kay Rinehart, Patrice-Andre Prud'homme, and Andrew Reid Huot, "Overwhelmed to Action: Digital Preservation Challenges at the under-Resourced Institution," *OCLC Systems & Services* 30:1 (February 2014): pp. 30-33.
[57] It is beyond the scope of this thesis to discuss the various ideas and methods of appraisal. An excellent and concise overview of appraisal theories throughout the history of archival studies can be found in Anne J. Gilliand's article "Archival Appraisal: Practicing on Shifting Sands" in *Archives and Recordkeeping: Theory into Practice*, ed. Caroline Brown, (London: Facet Publishing, 2014): pp. 31-62.

challenges to the traditional ideas of appraisal. There have been many attempts to find methods of efficiently and effectively appraising digital records including digital forensics[58] and automated appraisal.[59] Many of these methods, however, have proven difficult and are not widely adopted.[60] In their work on the Vancouver Olympic Games records at the City of Vancouver Archives, Courtney Mumma et al. suggest that appraisal is "predominantly a pragmatic exercise about "How much can archivists afford to keep?" rather than solely a theoretical exercise about "What documentary legacy do archivists want to pass to the future?"[61] The limitation of resources has a major impact on choosing records kept by the archives.[62] If a collection is well organized, has consistent file formats, and is large, it is much easier to select which records are worthy of preservation.[63]

This approach of managing digital records to effectively preserve them highlights the importance of records management. This means finding an efficient way to reduce the amount of digital records being accepted by the archive to facilitate appraisal. How to properly do this is a question that is open to much debate that does not appear to have a definitive solution and will generate different responses depending the individual. Taking all digital records into an archive might seem like an appropriate solution that will ensure that nothing is lost, but effectively managing this vast quantity of digital records is a difficult task that takes up large amounts of server space and staff resources spent on indexing and preservation. Alternatively, being highly

---

[58] Frederick B. Cohen, "Digital Diplomatics and Forensics," *Records and Information Management Journal*, 25:1 (March 2015): pp. 21-44; Christopher A. Lee, Matthew Kirschenbaum, Alexandra Chassanoff, Porter Olsen, and Kam Woods, "BitCurator: Tools and Techniques for Digital Forensics in Collecting Institutions," http://www.dlib.org/dlib/may12/lee/05lee.html, last accessed September 28, 2016.
[59] Gilliand, "Archival Appraisal," p. 49.
[60] Gilliand, "Archival Appraisal," p. 49.
[61] Courtney C. Mumma, Glenn Dingwall, and Sue Bigelow, "A First Look at the Acquisition and Appraisal of the 2010 Olympic and Paralympic Winter Games Fonds: or, SELECT * FROM VANOC_Records As Archives WHERE Value="true"," *Archivaria* 72 (Fall 2011): p. 110.
[62] Mumma et al., "A First Look," p. 110.
[63] Mumma et al., "A First Look," p. 121.

selective in the appraisal process may lead to acquiring valuable records that are easy to use and are worthy of preservation, it also takes significant effort on the part of staff to appraise so many digital records and also may lead to valuable digital records being lost. While automation may facilitate appraisal, it is by no means a replacement for human judgement and still requires the intervention of staff to function properly. In conclusion, there is no perfect solution to the challenges of appraising digital records. Each approach has its advantages and disadvantages. Some may work well in some situations and not in others. Therefore, institutions should select an appraisal approach based on their own needs and situations. There is no one size fits all approach.

To add further complexity, digital records change over time as they are migrated to new formats and storage media. This can lead to losses of functionalities and significant properties. A significant property is a characteristic of a record that is essential to the context and meaning of a record.[64] What constitutes a significant property is debated among members of the digital preservation community.[65] A significant property is difficult to identify due to the fluid nature of digital records[66] and the subjectivity of our views on value. To quote Angela Dappert and Adam Farquhar, "Significance is in the Eye of the Stakeholder."[67] A record does not have any value other than the one that we ascribe to it. Different groups of stakeholders may see different

---

[64] Geoffrey Yeo, "'Nothing Is the Same as Something Else': Significant Properties and Notions of Identity and Originality," *Archival Science* 10:2 (June 10, 2010): p. 87; Angela Dappert and Adam Farquhar, "Significance Is in the Eye of the Stakeholder," in *Research and Advanced Technology for Digital Libraries* eds. Maristella Agosti, José Borbinha, Sarantos Kapidakis, Christos Papatheodorou, Giannis Tsakonas (Berlin: Springer, 2009): p. 298.

[65] The term "significant property" is also not universally used in the digital preservation community. During my research, I have come across terms that refer to more or less the same thing. Examples include terms such as essence and significant characteristics.

[66] Digital records are fluid because the same digital record can exist on several computers and be displayed in slightly different ways depending on the configuration of these computers and the file format that the record is stored in. Are they still the same digital record? This question is open to debate as what constitutes an acceptable amount of loss is highly subjective. See Arthur Allison, James Currall, Michael Moss, and Susan Stuart, "Digital Identity Matters," *Journal of the American Society for Information Science and Technology* 56:4 (February 15, 2005): 364–72 for more on the fluid nature of digital records.

[67] Dappart and Farquhar, "Significance," p. 297.

properties that may be worth preserving in a record and at times these may conflict with each other. Geoffrey Yeo describes this subjectivity by explaining that records are "boundary objects." For Yeo, the interpretation of which properties of a record are significant varies from group to group.[68] For example, one group may consider the font of a textual document significant while another may put more importance on the actual words of the document. In sum, different communities ascribe significance to different properties of a record.

As a result, these disparate communities choose the appropriate digital preservation strategy that preserves the properties that they consider to be significant. Groups that rank content over appearance and functionalities will often see migration as a viable option for preservation. The risk of losing some of the digital record's appearance and original functionality is acceptable for the sake of preserving the content. For some communities, however, appearance and functionality are equal to content. A good example of this is video games. These digital records were designed to be played by the user and (for many communities, at least) this functionality is an integral aspect of experiencing these records. In other words, being able to play the game either in the exact way that the developer intended or as closely as possible is a significant property for many communities. Subsequently, digital preservation methods such as emulation are favoured over those that change the way the game was meant to be played.[69] Another example is preserving an entire operating environment to show what the creator of a digital record would have experienced.[70] In creating an emulation, the programmer selects what aspects of the digital record to preserve via emulation since emulation is far from a perfect method of preservation. It simply provides a simulation of what

---

[68] Yeo, "Nothing is the Same," pp. 97-99.
[69] Guttenbrunner et al., "Keeping the Game Alive," pp. 64-90.
[70] Carroll et al., "A Comprehensive Approach," pp. 61-92.

using the original digital record was like. The selection of an appropriate digital preservation method is determined by what is considered to be the significant properties of a digital record.

Both appraisal and significant properties demonstrate that the attribution of value in digital preservation is a subjective act that is influenced by a variety of factors. What may be considered valuable by some, may be seen as worthless to others. Different perceptions of value significantly affect decisions on which records are worth keeping and what significant properties are desirable. The same can be said for communities of individuals in their decisions about what is valuable. Each community has its own perception of what is valuable and will preserve digital records according to their own beliefs. This is manifest in the kinds of records that communities keep and the preservation strategies used to ensure that they are accessible. The choices that they make in digital preservation seek to preserve what they consider to be valuable. As we will see in Chapter Two, this community of users is what the Open Archival Information System (OAIS) calls the Designated Community.

Our perceptions of value affect our views on digital preservation and subsequently influence the choices that we make. There is no right or wrong answer to these questions. As long as the perceived value of the digital record is maintained throughout whatever preservation actions are performed, it can be said that the digital record in question has been preserved.

Conclusion

Digital preservation poses many challenges to information professionals. While many involve the rapid rate of technological advancement, it is important to keep in mind that this is a human problem rather than solely a technological one. It is through human action that machines and software advance and it is our agency that guides these tools in whatever task they perform.

The objects themselves do not act on their own but through human will. This is important to keep in mind as rapid improvements in technology have the tendency to generate hype and "hype hides history."[71] More exactly, hype hides the human agency involved in technological advancement. Nothing comes from nothing, and we should be mindful that there is always a human behind the machine.

How does this affect digital preservation? With each change in our society, our technology changes with us. We use these tools in different ways throughout our history. Records are tools created to satisfy certain requirements or perform specific tasks. It is the job of the archivist to ensure that this valuable context is not lost during preservation. In many instances, the challenges of digital preservation are ones of continuity. We need to preserve digital records that will more than likely change throughout their lifetime. The technology needed to render them will no longer be around so these records need to be migrated or changed in some way so that they can be read in some form by people. As a result, the paradoxical challenge of digital preservation is to preserve something that will change form. This often results in the selection of significant properties that are essential to keeping a record's context or "recordness" and using one of the digital preservation methods described above to preserve these properties. In response to these challenges, many standards and systems have been developed by the preservation community to facilitate preservation. These will be the focus of the next chapter.

---

[71] Mahoney, "The Histories of Computing(s)," p. 120.

Chapter Two

Standards and Digital Preservation Systems

As seen in the previous chapter, there are many challenges inherent in digital preservation. Through swift technological advances, older digital records become obsolete as the software and hardware that is required to render them is no longer available. This is assuming that a digital object is able to avoid having its bitstream degrade. In addition to technological challenges, digital preservation includes challenges that are more human in nature. Digital records can lose context and important meaning without adequate metadata to describe them. There is also the major issue of determining which digital records are worth preserving. Another significant challenge is determining what properties of a digital record are worthy of preservation. In essence, the challenge is to ascertain what we are trying to preserve.[1] Is the look of a digital record important, or is only the content? These are difficult questions to answer and there have been many standards and tools that have been developed in an attempt to address them.

The digital preservation community has created many systems and standards that attempt to respond to the challenges of digital preservation. As we have seen in chapter one, there were the Pittsburgh and UBC projects. Prior to 2000, the list of possible digital preservation options was short. There was an emphasis on media preservation with migration, emulation, and refreshing being developed. During the 2000s, digital preservation strategies such as migration were still mainstays but other methods were being developed including digital forensics, preservation metadata, standardized file formats, and data redundancy. Eventually, standards and guidelines were developed to aid in digital preservation.[2]

---

[1] See Yeo, "Nothing is the Same," and Guttenbrunner et al., "Keeping the Game Alive."
[2] Harvey, *Preserving Digital Materials*, pp. 101-105.

The response of archives to digital preservation has varied. In her 2013 study on how many institutions have digital preservation policies, Madeline Sheldon found that only 33 of the 99 organizations she surveyed had developed a digital preservation policy between 2008 and 2013.[3] Texas Tech University began a program to keep electronic theses and dissertations (ETDs) in 2005. During the course of operations, the program had suffered numerous instances of data loss and other preservation issues such as human error and corrupted backups. While they were still able to continue their program, it required much work on the part of staff.[4] LAC is another institution that has faced many challenges in their digital preservation efforts. Despite declaring that digital records will be the "format of choice" for the government by 2017, LAC does not have a digital strategy in place. As a result, it is unable to effectively and efficiently handle large quantities of digital records.[5] In addition, LAC spent $15.4 million on a trusted digital repository (TDR) between 2006 and 2011. While this TDR was deemed operational in July 2011, it was shut down without official documentation of the reasoning in November 2012.[6] It was estimated in 2013 that the Government of Canada had 14 PB worth of unmanaged emails in its archives and email systems.[7]

As can be seen from these examples, the responses to digital preservation have met with mixed results. While it would take volumes to go into detail about all of the responses to digital

---

[3] Madeline Sheldon, *Analysis of Current Digital Preservation Policies: Archives, Libraries, and Museums*, (2013), http://www.digitalpreservation.gov/documents/Analysis%20of%20Current%20Digital%20Preservation%20Policies.pdf, last accessed September 28, 2016.

[4] Joy Perrin, Heidi Winkler, Le Yang, "Digital Preservation Challenges with an ETD Collection – A Case Study at Texas Tech University," *The Journal of Academic Librarianship* 41 (2015): pp. 98-104.

[5] Auditor General of Canada, "Chapter 7 Documentary Heritage of the Government of Canada – Library and Archives Canada," Fall 2014, pp. 8-10.

[6] Auditor General of Canada, pp. 11-12. For more information on LAC's TDR, see Greg Bak, "Trusted By Whom? TDRs, Standards Culture, and the Matter of Trust," *Archival Science* 16:6 (Fall 2016).

[7] Kristina Lillico, "Preparing for the Digital World in the Federal Government," *Navigating the Digital Future IM Conference*, (November 27, 2013): p. 9, http://www.im.gov.ab.ca/documents/conference/2.1_Preparing_for_the_Digital_World_-_KLillico.pdf, last accessed September 28, 2016..

preservation, in this chapter I will focus on specific standards that have been influential in the digital preservation community and some of the digital preservation systems that arose from them. In particular, I will examine OAIS and PREMIS and explore their influence in the creation of several digital preservation systems such as Archivematica and Preservica. The standards and guidelines that have been developed by the digital preservation community are influential in how these two digital preservation systems operate as these systems are built around them. I will start with the standards and then move on to these systems.

Open Archival Information System (OAIS)

One of the most influential standards in digital preservation is the Reference Model for an Open Archival Information System (OAIS), a standard whose development was led by the Consultative Committee for Space Data Systems (CCSDS). The CCSDS was created in 1982 as a forum for the discussion of problems in the development of space data systems. It is made up of many agencies from various nations across the globe.[8] Prior to the development of OAIS, there were no consistent standards on digital preservation. Beginning in the 1950s, organizations began to rely on computers to store their data and throughout the 1960s and 1970s there was an increasing awareness of digital preservation issues. Through the 1980s and early 1990s there were efforts on the parts of archivists and others who handled digital data to address the challenges of digital preservation.[9] At the request of the International Organization for Standardization (ISO) the CCSDS agreed to join them to develop a consistent digital preservation standard in 1990. The work done by the CCSDS was to be reviewed by ISO and

---

[8] The Consultative Committee for Space Data Systems (CCSDS), "About CCSDS," http://public.ccsds.org/about/default.aspx, last accessed September 28, 2016.

[9] Christopher Lee, "Open Archival Information System (OAIS) Reference Model," *Encyclopedia of Library and Information Studies, Third Edition* (2010): pp. 4020-4021.

possibly become an international standard.[10] In 1995, CCSDS hosted an international workshop that developed the basic framework for OAIS. Draft versions of OAIS were released in May 1997 and May 1999 and in June 2000, ISO published it as a draft ISO standard. It was finally approved in January 2002 as ISO Standard 14721.[11]

OAIS is "open" in the sense that it is developed in an open forum that allows user participation.[12] The goal of OAIS is to ensure long term digital preservation and access to digital records for the designated community. This can mean preservation indefinitely or for only a short period of time.[13] Finally it should be noted that OAIS is a reference model that provides guidelines for organizations to follow rather than formal rules.[14] This is important as OAIS does not provide a specific blueprint for an organization to follow. What it does instead is to provide a general framework within which digital preservationists can work.

OAIS has become a popular reference model for digital preservation professionals across the globe. In a survey of 48 institutions across 13 nations, 80% of respondents claimed that their repository either fully or partly conformed to OAIS.[15] Paul Laughton also found that there was high degree of compliance with OAIS. From the 26 responses that he received from various institutions, he found that the majority of them scored over 61 points out of a possible 92 in their OAIS compliance. Only four institutions scored their OAIS compliance below 50 points.[16] What is clear from these examples is that OAIS is influential in the digital preservation

---

[10] Brian F. Lavoie, *The Open Archival Information System Reference Model: Introductory Guide* (January 2004): pp. 2-3.
[11] Lavoie, *The Open Archival Information System*, pp. 2-3; Lee, "Open Archival Information System," pp. 4021-4023.
[12] CCSDS, *Reference Model for an Open Archival Information System (OAIS)*, June 2012, section 1, pp. 1.
[13] CCSDS, *OAIS*, section 1, pp. 1.
[14] CCSDS, *OAIS*, section 1, pp. 1-2.
[15] Jaqueline Spence, "Preserving the Cultural Heritage: An investigation into the Feasibility of the OAIS Model for Application in Small Organisations" *Aslib Proceedings: New Information Perspectives* 58:6 (2006), p. 514.
[16] Paul Laughton, "OAIS Functional Model Conformance Test: A Proposed Measurement" *Program: Electronic Library and Information Systems* 46:3 (2012), p. 319.

community. As these surveys suggest, many organizations have used OAIS in some shape or form in developing their digital preservation repositories. But why is OAIS so popular? How do OAIS guidelines facilitate digital preservation?

An important aspect of OAIS is that it is divided into functional entities. There are six functional entities within OAIS: Ingest, Archival Storage, Data Management, Administration, Preservation Planning, and Access.[17] The Ingest Functional Entity involves receiving submission information packages and preparing them for the Archival Storage Functional Entity which stores and maintains the Archival Information Packages. The Data Management Functional Entity maintains the data that is stored within the OAIS while the Administration Functional Entity provides management of the archive and negotiates submission agreements with producers. The Preservation Planning Functional Entity monitors the OIAS to make sure that it is accessible to the designated community while the Access Functional Entity supports consumers.[18] In addition, there are many groups of users that interact with the OAIS. Producers create the information that will be stored within the archive while managers set the policy of the archive. It is important to note that while managers set the archival policies, they are not the ones who run the daily operations of the repository. This is done by the archivists. Consumers are people who use the information within the archive and they are likely members of the fourth group within OAIS, the designated community.[19] The designated community is a specific group of people that may include both users and consumers who are the main focus of the archive. The designated community posseses what OAIS terms a "knowledge base" which allows it to understand the information stored within the archive. As we will see, this knowledge base

[17] CCSDS, *OAIS*, section 4, pp. 1-3.
[18] CCSDS, *OAIS*, section 4, pp. 1-3.
[19] CCSDS, *OAIS*, section 2, pp. 2-3.

changes over time, which results in changes in the representation information that the archive needs to attach to information objects.[20]

Within the OAIS reference model, an information object is composed of a data object and its representation information. A data object is composed of one or more digital objects such as a DOCX or CSV file. Representation information is the data that is bundled with a data object which makes it intelligible to the designated community. It is divided into three kinds, structural, semantic, and other. Structural representation information is information about the file format that the data object is in while semantic representation information describes the meaning of the object and ontologies. Other representation information is information that is neither structural or semantic. Examples of this kind of representation information include encryptions, software, written instructions, and algorithms. Information that describes the relationship between structural and semantic information would also be considered other representation information.[21]

As a result, it is possible to have a complex web of representation information that includes its own information objects for a single information object in an archive. How much and what kinds of representation information to include in an information object is determined by the knowledge base of the designated community. A knowledge base is the level of knowledge that the designated community has in interpreting information objects. Understanding the designated community's knowledge base is crucial for an archive because it will determine how much representation information is necessary for the information object to be usable. A designated community's knowledge base can change over time and this can lead to a revision in the amount of representation information that is attached to data objects.[22]

---

[20] CCSDS, *OAIS*, section 2, p. 3.
[21] CCSDS, *OAIS*, section 1, p. 13; section 2, pp. 3-5; section 4, pp. 20-23.
[22] CCSDS, *OAIS*, section 2, pp. 3-5; section 4, pp. 20-23.

The interplay between the designated community's knowledge base and representation information is an important factor in OAIS. With OAIS, the archive's main responsibility is to ensure that records are accessible and understood by the designated community. As a result, this dictates how much representation information needs to be included within the information object. In many instances, archivists describe their holdings and shape their meaning to accommodate the designated community. Archivists decide how much representation information to add to the data object so that it can be an information object that is understandable to the designated community. This can also change the way an archive describes its data objects due to the shifting knowledge base of the designated community.

Representation information is complex and often involves its own information objects and therefore its own sets of representation information. There will be additional sets of representation information until the original information object and its representation information is understandable to the designated community. As the designated community's knowledge base changes, so too does this representation information network.[23] The resulting complex network of representation information and the information object that it describes are stored within the information package.

The information package is what contains information within an archive. More specifically, there are two kinds of information contained within an information package. These are content information and preservation description information (PDI). Content information is the information object (i.e. the data object and its representation information) that is the target of preservation within the archive. PDI, however, is the data that describes the content information so that it is identifiable and understandable. There are five parts to PDI, provenance (the history

[23] CCSDS, *OAIS*, section 3, pp. 23-25.

of the content information), reference (unique identifier), context (the content information's context), fixity (protects content information), and access rights.[24] In addition to this information the information package is further encased within packaging information which describes how the content information relates to the PDI. Descriptive information is then finally added to the information package so that it can easily be identified.[25]

In OAIS, there are three kinds of information packages. The first kind is the submission information package (SIP). This is the information package that is produced by the producer and is submitted to the archive. It contains content information and some PDI and is used by the archive to create the archival information package (AIP). An AIP can be made up of one or more SIPs or even part of a single SIP. An AIP also contains a full set of PDI in addition to the content information. In response to a research request from consumers, all or part of the AIP is used to generate a dissemination information package (DIP) that can be used by researchers. The DIP will contain the content information but might not contain the full set of PDI found within the AIP.[26] These different information packages allows information to flow into and out of an archive. When the producers contribute information to the archive, it is submitted in a SIP which is then turned into an AIP by the archive.[27] When a consumer requests information, a DIP is sent out to the consumer.

OAIS provides a guideline for institutions to follow in setting up a digital preservation system. It supplies a framework for other organizations to create their own digital archives and supplies a common vocabulary that facilitates collaboration. It is crucial to remember, however, that OAIS is a guideline and not meant to provide prescriptive rules for an implementation of a

[24] CCSDS, *OAIS*, section 2, pp. 5-7.
[25] CCSDS, *OAIS*, section 2, p. 7.
[26] CCSDS, *OAIS*, section 2, pp. 7-8.
[27] CCSDS, *OAIS*, section 4, p. 52.

digital archive. It is meant to be a general guide and not a set of procedures.[28] In Natascha Schumann and Astrid Recker's article, "De-Mystifying OAIS Compliance: Benefits and Challenges of Mapping the OAIS Reference Model to the GESIS Data Archive," they examine the myth behind OAIS compliance. They suggest that complying "with the OAIS model means complying with a set of very abstract requirements which themselves need interpretation, translation, and concretization if they are to be useful."[29] In other words, OAIS compliance is up to interpretation and is otherwise "utterly meaningless"[30] with an institution having varying degrees of "OAIS compliance."[31]

The abstract character of OAIS leads to many challenges that are not covered in the reference model. One such challenge is the important role that information management plays in digital preservation. Adrian Cunningham believes that OAIS does not effectively address the challenge of managing the vast quantities of digital records prior to ingest into the archive. For Cunningham, OAIS does not have any guidance for creating and retrieving "reliable records that can serve as evidence of decisions and activities among the mountains of what are often dynamic, anarchic, and unmanaged data"[32] that are created by individuals and organizations.[33] As we have seen in chapter one, effectively managing the immense quantity of digital records is a significant challenge in digital preservation. The challenge of managing digital records and attributing value to those that are deemed archival is not addressed in OAIS. The reference model only describes the ingest process and not the management of information prior to the

---

[28] Natascha Schumann and Astrid Recker, "De-Mystifying OAIS Compliance: Benefits and Challenges of Mapping the OAIS Reference Model to the GESIS Data Archive," *IASSIST Quarterly* 36:2 (2012): p. 6.
[29] Schumann and Recker, "De-Mystifying OAIS," p. 7.
[30] Schumann and Recker, "De-Mystifying OAIS," p. 10.
[31] Schumann and Recker, "De-Mystifying OAIS," p. 7.
[32] Adrian Cunningham, "Digital Curation/Digital Archiving: A View from the National Archives of Australia," *The American Archivist* 71:2 (2008): pp. 535.
[33] Cunningham, "Digital Curation," pp. 533-535.

formation of the SIP. While this may be beyond the scope of OAIS, it does not address this crucial aspect of digital preservation.

In a similar vein, the concept of a designated community is also problematic. As discussed above, a designated community consists of the users that the archives preserves records for. A designated community also posseses a knowledge base from which the archive can determine how much representation information to add to the data object to effectively change it into a meaningful information object. Some designated communities, such as space scientists, have a relatively homogenous knowledge base which makes determining the appropriate amount of representation information significantly easier. But what happens when the designated community is made up of users who do not have a consistent knowledge base? This is the problem that Jerome McDonough had in applying OAIS to the Preserving Virtual Worlds (PVW) project. The goal of this project was to develop standards for content representations and metadata to preserve computer games in many repositories.[34] One of the main problems that McDonough and the PVW team had in implementing OAIS was that the repositories that they were working with were university campuses whose mandates required them to serve the entire university including staff, students, and visitors. This designated community has no consistent knowledge base which in turn produces the challenge of determining how much representation to include with a data object to make it an intelligible information object for the designated community.[35]

The PVW project found it extremely difficult to gauge the appropriate level of representation information to add to a data object for the user to completely understand the data.

---

[34] Jerome P. McDonough, "'Knee-Deep in the Data:' Practical Problems in Applying the OAIS Reference Model to the Preservation of Computer Games," *45th Hawaii International Conference on System Science* (2012), p. 1625.
[35] McDonough, "Knee-Deep in the Data," pp. 1629-1630.

In one example, McDonough describes how the amount of representation needed that a person with little computer science background. McDonough comments that "Nothing less than a small library of works necessary to provide a basic education in computer science would suffice."[36] As can be seen with the PVW project, the assumption of OAIS that the designated community would have a homogenous knowledge base is highly problematic.[37] Other institutions have also reported on the challenges of describing for a designated community. In their work at the Inter-University Consortium for Political and Social Research (ICPSR), Mary Vardigan and Cole Whiteman noted that in their application of OAIS, they will have to modify their descriptive practices to include the non-expert researchers that the ICPSR works with.[38] Determining the appropriate levels of information required to sufficiently understand a digital record is a difficult task that OAIS only touches upon. It is one of many critical aspects of digital preservation that OAIS does not sufficiently address.

Despite its widespread acceptance, the abstract nature of OAIS leads to many challenges. This leads to certain aspects of OAIS to be up to interpretation. This open nature leads to many of the challenges described above. It should be kept in mind, however, that OAIS is meant to be a guide that can be incorporated into a digital preservation program. It is not a specific digital preservation program that can be applied in any institution. OAIS contains many pieces of good information on how to preserve digital content, but it needs to be used with other standards and practices to be fully effective. One of these standards is the PREservation Metadata Implementation Strategies (PREMIS).

---

[36] McDonough, "Knee-Deep in the Data," p. 1630.
[37] McDonough, "Knee-Deep in the Data," p. 1629.
[38] Mary Vardigan and Cole Whiteman, "ICPSR Meets OAIS: Applying the OAIS Reference Model to the Social Science Archive Context," *Archival Science* 7:1 (2007): pp. 77-78.

PREMIS

PREMIS is a metadata standard that focuses on preservation metadata. The need for a consistent standard on preservation metadata was first articulated in 1996 by John Garret and Donald Waters when they encouraged the archival community to create a system to safeguard digital records from destruction.[39] For Garret and Waters, they believed that "a safety-net is needed to ensure that digital information objects with long-term cultural and intellectual value survive the expressions of stakeholder interest with their integrity intact."[40] Garret and Waters articulated the need to preserve the integrity of digital records, which required information on provenance and context. These ideas influenced the preservation description information that is present in OAIS.[41]

These ideas also influenced others to begin work on a solution. In 2001, the Online Computer Library Centre (OCLC) and Research Libraries Group (RLG) established the Preservation Metadata Framework working group to explore the types of information that should be associated with digital records. Their 2002 report, *A Metadata Framework to Support the Preservation of Digital Objects*, outlined many metadata elements that may be used in digital preservation. This report focused on OAIS and sought to expand its metadata components models. Because of the abstract nature of OAIS,[42] additional work was needed to make these

---

[39] Devan Ray Donaldson and Elizabeth Yakel, "Secondary Adoption of Technology Standards: The Case of PREMIS," *Archival Science* 13:1 (May 11, 2012): pp. 59-60; John Garret and Donald Waters, *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information*, Council on Library and Information Resources (Washington D.C., 1996), http://www.clir.org/pubs/reports/pub63watersgarrett.pdf, pp. 19-20, last accessed September 28, 2016.
[40] Garret and Waters, *Preserving Digital Information*, p. 20.
[41] Donaldson and Yakel, "Secondary Adoption," p. 60.
[42] Online Computer Library Center (OCLC) and Research Libraries Group (RLG), *A Metadata Framework to Support the Preservation of Digital Objects*, OCLC Online Computer Library Center (Dublin, Ohio: 2002): http://www.oclc.org/content/dam/research/activities/pmwg/pm_framework.pdf, p. 47, last accessed September 28, 2016.

prototype metadata elements usable in a repository.[43] Therefore, the PREMIS working group was established in 2003 to create a set of metadata requirements for digital preservation.[44] The working group was also tasked with providing guidance and suggesting best practice for managing, using, and creating metadata for digital preservation.[45] The initial project lasted until 2005 when the working group published its report, *PREMIS Data Dictionary for Preservation Metadata*. The report included a data dictionary in addition to information about preservation metadata.[46]

After the release of *PREMIS Data Dictionary* in 2005 and through the sponsorship of the Library of Congress, the PREMIS Maintenance Activity was created to maintain the data dictionary and raise awareness about preservation metadata and similar topics. The Maintenance Activity also provides training resources, an XML schema, and a Web home for the data dictionary.[47] Many other working groups were created by the Maintenance Activity to aid in the promotion and support of PREMIS. An Editorial Committee was created to check for errors in the data dictionary and XML schema. The PREMIS Implementers Group (PIG) was created to commission studies on preservation metadata topics and to provide tutorials on PREMIS. They also operate discussion lists and a wiki.[48] User participation is important in PREMIS as users provide feedback and report errors to PIG and the Maintenance Activity.[49] When enough

---

[43] Library of Congress (LOC), *PREMIS Version 3.0*, http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf, p. 1, last accessed September 28, 2016.
[44] Priscilla Caplan, and Rebecca Guenther, "Practical Preservation: The PREMIS Experience," *Library Trends* 54:1 (Summer 2005): pp. 111-112.
[45] LOC, *PREMIS Version 3.0*, p. 1.
[46] Priscilla Caplan, "Understanding PREMIS," http://www.loc.gov/standards/premis/understanding-premis.pdf, (February 2009): p. 4, last accessed September 28, 2016.
[47] LOC, *PREMIS Version 3.0*, p. 3.
[48] LOC, *PREMIS Version 3.0*, p. 3.
[49] User participation is also important in the development of standards such as OAIS. For more information regarding on how standards are created please see Alan Bell, "Standards and Standards Culture: Understanding the Nature and Criticisms of Standardisation," in *Comma* (Paris, 2001): pp. 29-32.

documentation has been gathered on any particular issue, it is reviewed by the Editorial Committee.[50] Since its release in 2005, there have been many updates to PREMIS. In March 2008, a second version was released[51] and a third version was released in June 2015.[52]

The goal of the PREMIS Data Dictionary is to provide "a comprehensive, practical resource for implementing metadata in digital preservation systems."[53] For PREMIS, metadata is defined as the information that an institution uses to effect digital preservation. PREMIS' metadata schema spans many categories that are traditionally considered to be separate types of metadata such as technical, structural, rights, and administrative.[54] PREMIS pays particular attention to digital records' provenance and in version 3.0 there is a focus on implementing metadata throughout a record's life-cycle.[55] Finally, PREMIS is "implementation independent." This means that the PREMIS Data Dictionary is designed to work regardless of how a preservation system was created.[56] PREMIS' implementation independent approach is similar to OAIS in that both of these standards do not provide specifics on how to implement them. Rather, it is up to the user to determine how PREMIS and OAIS could be used by an organization.

PREMIS defines elements, which are called semantic units, which describe the properties of and relationships between entities within a digital preservation system.[57] They are described using metadata elements such as entities. Entities are the individual units within a digital

---

[50] LOC, *PREMIS Version 3.0*, pp. 3-4.
[51] Caplan, "Practical Preservation," p. 4.
[52] Library of Congress, PREMIS Website, http://www.loc.gov/standards/premis/v3/index.html, last accessed September 28, 2016.
[53] LOC, *PREMIS Version 3.0*, p. 1.
[54] LOC, *PREMIS Version 3.0*, p. 2. For more information on the specific types of metadata, please see Remesh et al. pp. 195-196.
[55] LOC, *PREMIS Version 3.0*, p. 2.
[56] LOC, *PREMIS Version 3.0*, p. 3.
[57] LOC, *PREMIS Version 3.0*, pp. 6-7.

preservation unit. There are many kinds of entities including objects (a discrete unit of digital information), events (an action that involves at least one agent or object), agents (software/organizations/individuals who are associated with an object's events and/or rights), and rights (permissions regarding an agent and/or an object).[58] In short, PREMIS outlines the semantic units (properties) that are used to describe entities.

Entities often contain many sub-types that are useful for understanding the semantic units within a digital preservation system. Version 3.0 of PREMIS divides objects into four categories: intellectual entities, representations, files, and bitstreams.[59] An intellectual entity is a unique artistic or intellectual creation.[60] An example being a photograph, book, video, or a web page. Prior to version 3.0, PREMIS did not include intellectual entities in its data dictionary as they were not considered to be required for digital preservation and were already well serviced with other descriptive metadata.[61] They were introduced in version 3.0 because many institutions needed a way to describe the content that was common across many representations of the same intellectual entity.[62]

Representations, PREMIS' equivalent to OAIS' information objects, are crucial for a digital preservation system as they are needed to display intellectual entities. An example of a representation is a TIFF image of a building or an MKV file of a movie. Without a representation, an intellectual entity would only remain an abstract concept.[63] Consider this example: I take a photograph of a boat and I save it in various formats. The intellectual entity is the image of the boat while the representations are the files that show this image. I still

[58] LOC, *PREMIS Version 3.0*, p. 7.
[59] Angela Dappert, Sébastien Peyard, Carol C.H. Chou, and Janet Delve, "Describing and Preserving Digital Object Environments," *New Review of Information Networking* 18:2 (2013): pp. 109-110.
[60] LOC, *PREMIS Version 3.0*, p. 9.
[61] LOC, *PREMIS Version 3.0*, p. 9.
[62] LOC, *PREMIS Version 3.0*, p. 11.
[63] LOC, *PREMIS Version 3.0*, p. 12.

experience the same image regardless of the file format that I view the image in. Furthermore, a representation may be composed of one or more files. Files are the discrete digital objects that make up a representation.[64] Examples include JPG files or TIFF images. Finally, each file is composed of a series of bits called a bitstream.[65] In the example of the boat given above, each representation is composed of one file and its bitstream. An example of a representation having many files is a web page where there is the HTML file as well other files.

These separate kinds of objects share different types of relationships with each other. In the PREMIS Data Dictionary, many of the relationships between the different object types are either structural or derivation relationships. Structural relationships describe how the different parts of an object relate to each other. This is, for example, how an intellectual entity relates to its representation and how this representation relates to its files and vice versa.[66] The image of the boat is the intellectual entity which has a structural relationship with the JPG file of an image (a representation). Derivation relationships are the outcomes of an object's transformation or replication.[67] If I have a representation in a JPG file and I migrate it into a TIFF file, this new TIFF file will have a derivation relationship with the JPG file. These two relationships are the most common and it is noted in the data dictionary that there "is no way to model all possible structural and derivation information."[68] It is up to the repository to determine which metadata is required to document these relationships. PREMIS just provides the basic framework within which to work and acknowledges that implementers may use other metadata schemas when needed.[69]

---

[64] LOC, *PREMIS Version 3.0*, pp. 12-13.
[65] LOC, *PREMIS Version 3.0*, p. 13.
[66] LOC, *PREMIS Version 3.0*, p. 19.
[67] LOC, *PREMIS Version 3.0*, p. 19.
[68] LOC, *PREMIS Version 3.0*, p. 19.
[69] LOC, *PREMIS Version 3.0*, pp. 19-20.

Version 3.0 of PREMIS also introduces the concept of environments as object entities.[70] Prior to version 3.0, there was an environment container with an object's metadata but now environments are their own standalone objects. This was done to record the important provenance information relating to how digital records operated within their original operating environment.[71] There are many digital records and artifacts that can fall within the environments category including computer hardware, operating systems, and software applications. With the PREMIS Data Dictionary, each one of these environment entities is described separately and linked to the non-environment object (the content object). This ensures that the relationship between file, software, and hardware is clearly described to researchers and to preserve a digital record's original operating context.[72] Important contextual metadata that is captured by the data dictionary includes the purpose of the environments and how they operate.[73]

With the inclusion of environment entities, PREMIS has created a special relationship category to describe the interplay between software and the objects that they run. Dependency relationships document how environment entities interact with each other.[74] An example of a dependency relationship would be a computer game that requires Windows to run. This computer game is an application that is dependent on the Windows operating system to function. Therefore, the game has a dependency relationship with Windows. By having this relationship described, PREMIS is able to able to link hardware and software to digital records.[75]

These added features in PREMIS version 3.0 are important for digital preservation because, as we have seen in chapter one, digital records are dependent on technology to be

[70] For additional information on why environments were added PREMIS, see Dappert et al., "Describing and Preserving," pp. pp. 106-173.
[71] LOC, *PREMIS Version 3.0*, p. 251.
[72] LOC, *PREMIS Version 3.0*, pp. 251-253.
[73] LOC, *PREMIS Version 3.0*, pp. 253-254.
[74] LOC, *PREMIS Version 3.0*, p. 20.
[75] LOC, *PREMIS Version 3.0*, p. 20.

properly rendered and used by people. The relationship between technology and digital records is important for all preservation strategies as it provides vital contextual information to how a digital record functioned. For example, knowing a digital record's file format and operating environment will help us render it or recreate the original operating environment via emulation. Good examples of this are the emulation of video games and the Salman Rushdie emulation at Emory University.[76] In both of these examples, recreating the digital record's operating environment was crucial because it provided valuable contextual information as to how the record was used. In turn, this contextual information ensures that digital record's integrity remains intact.

Another entity that is used in PREMIS are events. Events are the actions that occur on an object. Examples include migration, deaccessioning, or even a simple integrity check.[77] All events have an outcome and some have outputs. Outcomes are whether the event was a success or a failure while an output is the creation of a new object (such as the end result of migration).[78] This documentation of the actions performed on a record is valuable contextual information for archivists because it provides information on the record's provenance. We know when it changed custody and what preservation actions were performed throughout the records lifetime. Agents are another entity within the PREMIS Data Dictionary. What constitutes an agent can be very broad as it can range anywhere from an individual to an organization and even software. An agent is someone or something that affects an object.[79] Finally, the last entity is rights. The

---

[76] For video game emulation, see Guttenbrunner et al, "Keeping the Game Alive." For the Rushdie Emulation, see Carroll et al., "A Comprehensive Approach."
[77] LOC, *PREMIS Version 3.0*, p. 15.
[78] LOC, *PREMIS Version 3.0*, p. 16.
[79] LOC, *PREMIS Version 3.0*, p. 16. There can be some confusion as to why software can be both an agent and an environment. The point of the software as environment entity is to further explain the software as an agent entity.

rights entity describes the access, copying, and modification permissions that are placed on an object. This includes copyright and intellectual property.[80]

These different kinds of entities have many types of relationships with each other. PREMIS is used to describe these relationships and properties. The interplay between entities within the PREMIS Data Dictionary describe valuable information that not only can be used to keep digital records within context and illuminate their provenance, but also make them renderable. In many ways, PREMIS seeks to address the challenge of digital records lacking appropriate metadata. As we have seen in chapter one, the need to collect sufficient amounts of metadata is often ignored[81] and there is little consensus between metadata standards.[82] PREMIS attempts to rectify this by outlining the various entities that organizations need to describe in order to successfully preserve a digital record.[83] Like OAIS, PREMIS provides a common framework within which a repository can develop its own metadata policies and procedures. While implementations of PREMIS may share a common vocabulary, there may be several differences depending on the implementing organization's needs.[84]

The result of this general nature is that organizations need to modify PREMIS and/or their own institutions for the standard to function. In their article, "Implementing PREMIS: A Case Study of the Florida Digital Archive," Devan Ray Donaldson and Paul Conway describe some of the challenges of starting a PREMIS implementation. In many instances, the Florida Digital Archive needed to modify its organizational practices and PREMIS to successfully use

---

[80] LOC, *PREMIS Version 3.0*, p. 17. Another good source for the rights entity is Karen Coyle, *Rights in the PREMIS Data Model*, http://www.loc.gov/standards/premis/Rights-in-the-PREMIS-Data-Model.pdf, last accessed September 28, 2016.
[81] Breytenbach and Groenewald, "The Use of Metadata," p. 237.
[82] Ramesh et al., "Metadata Diversity," pp. 196-198; Wilson, "How Much is Enough," pp. 212-214.
[83] LOC, *PREMIS Version 3.0*, p. 24.
[84] LOC, *PREMIS Version 3.0*, p. 24.

the data dictionary.[85] Elizabeth Yakel and Devan Ray Donaldson also examined the challenges of adopting PREMIS. In their article, "Secondary Adoption of Technology Standards: The Case of PREMIS," the authors found many common themes among the organizations interviewed. These include strong managerial support for PREMIS as well as available resources to implement the data dictionary. Many of these institutions also work in environments that encourage experimentation, which led to modifying PREMIS and combining it with other metadata standards.[86] The need to combine PREMIS with other metadata standards is crucial due to the fact that PREMIS exclusively focuses on the metadata that is needed for digital preservation. It does not include metadata options for access, archival description, and organization.[87]

These examples also highlight the need for considerable organizational support for the implementation of PREMIS. In the study by Donaldson and Yakel, nearly all of the institutions interviewed were universities and national archives or libraries.[88] These institutions often have the budget to acquire a wide assortment of technical expertise. The Florida Digital Archive also had both the human and financial resources to experiment with PREMIS.[89] This leads to the question; how can smaller organizations successfully manage metadata? While it is far beyond the scope of this thesis to suggest methods of harvesting metadata and the specific ways to modify PREMIS or other metadata standards, I feel that this is a significant challenge to many institutions that do not have many technical and financial resources.

---

[85] Devan Ray Donaldson and Paul Conway, "Implementing PREMIS: A Case Study of the Florida Digital Archive," *Library Hi Tech* 28:2 (2010): pp. 283-286.
[86] Donaldson and Yakel, "Secondary Adoption," pp. 76-79.
[87] Wilson, "How Much is Enough," pp. 213-214.
[88] Donaldson and Yakel, "Secondary Adoption," pp. 64-65.
[89] Donaldson and Conway, "Implementing PREMIS," pp.277-278.

The PREMIS Data Dictionary is a valuable tool for digital preservation as it provides a common vocabulary for describing preservation metadata. Since its inception PREMIS has become the most widely used standard among the digital preservation community in regard to preservation metadata.[90] It is considered by Brown to be fast "becoming the *de facto* international standard for preservation metadata."[91] PREMIS has been widely influential within the digital preservation community. According to Alan Bell, influence and widespread adoption in a community are crucial for changing best practices into standards.[92] This also means that the "best standard" is not always guaranteed to become a profession's *de facto* standard. Rather, it is through widespread adoption that something becomes a standard.[93] This often leads to standards having flaws and as we have seen from the examples above, PREMIS, much like OAIS, is far from perfect. PREMIS is often written in abstract language and requires the implementing repository to modify the data dictionary to suit its needs. PREMIS also only focuses on preservation metadata and needs to be paired with other metadata standards to fully describe a digital record. Finally, the technical expertise required to implement a metadata program may be out of the reach of some institutions. Much like OAIS, PREMIS is useful but does require some work on the part of the implementer to function in an actual workplace.

Both OAIS and PREMIS have been influential in the digital preservation community and two such products where their influence can be felt are Artefactual's Archivematica and Tessella's Preservica.

Archivematica

---

[90] Wilson, "How Much is Enough," p. 212.
[91] Brown, *Practical Digital Preservation*, p. 167.
[92] Bell, "Standards," p. 36.
[93] Bell, "Standards," p. 27.

Archivematica is an open source digital preservation system that was first developed by Artefactual Systems Inc. in 2009. It began as the back-end digital preservation component of Access to Memory (AtoM), another one of Artefactual's products. Initially called Qubit-OAIS, Artefactual eventually renamed the system Archivematica.[94] In June 2007, Kevin Bradley, Junran Lei, and Chris Blackall released a report for the UNESCO Memory of the World Programme Sub-Committee on Technology that explained the need for an open source digital preservation system that was compliant with OAIS. The report also described the need to make this preservation system affordable and widely available.[95]

Archivematica is released for free under a GNU (GNU's Not Unix)[96] Affero General Public License[97] with paid service plans available.[98] Archivematica's documentation is also released for free under a Creative Commons Attribution-ShareAlike 4.0 International License.[99] As a result, organizations can use Archivematica however they see fit. The goal of Archivematica is to provide a comprehensive and OAIS compliant digital preservation system to information professionals who may have limited technical experience and/or resources.[100] Other

[94] Peter Van Garderen, P. Jordan, T. Hooten, C. Mumma, and E. McLellan, "The Archivematica Project," http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/VC_Van_Garderen_et_al_26_Workshop1.pdf, p. 3, last accessed September 28, 2016.
[95] Van Garderen, et al., "The Archivematica Project," pp. 2-3.
[96] GNU is the name of an operating system.
[97] Artefactual, "What is Archivematica?" https://www.archivematica.org/en/docs/archivematica-1.4/user-manual/overview/intro/#intro, last accessed September 28, 2016. For information regarding a GNU Affero General Public License, please see their website at http://www.gnu.org/licenses/agpl-3.0.en.html, last accessed September 28, 2016.
[98] Artefactual, "Services," https://www.artefactual.com/services/, last accessed September 28, 2016.
[99] Artefactual, "What is Archivematica?" https://www.archivematica.org/en/docs/archivematica-1.4/user-manual/overview/intro/#intro, last accessed September 28, 2016. For information regarding the Creative Commons License, please see http://creativecommons.org/licenses/by-sa/4.0/, last accessed September 28, 2016.
[100] Artefactual, "What is Archivematica?" https://www.archivematica.org/en/docs/archivematica-1.4/user-manual/overview/intro/#intro, last accessed September 28, 2016.

standards that are used within Archivematica include Dublin Core, METS, Library of Congress Bagit Specification, PREMIS, and many others.[101]

To achieve compliance with so many standards, Archivematica is structured around the use of micro-services. Micro-services are discrete open source programs that are bundled within Archivematica. Each of these micro-services handle specific tasks within Archivematica's workflows. For example, ClamAV scans for viruses and malware while Format Identification for Digital Objects (FIDO) identifies the file type during the transfer process.[102] In addition to these micro-services, Archivematica also has its own Format Policy Registry (FPR). The FPR is a database that allows Archivematica to identify format policies for specific file formats. A format policy identifies the appropriate course of action for preserving a file format. For example, if a word processing document is identified, the FPR will contain the appropriate set of instructions to preserve this file format.[103] Once identified, Archivematica will then normalize the formats of the digital object to a preservation format according to the FPR.[104] The combination of the FPR and open source micro-services allows Archivematica to preserve various digital records. The FPR will identify what preservation action needs to be undertaken while the micro-services perform the tasks of preserving the digital objects. All of this is available in a single package that is freely available to download from the web. This free and open accessibility has made Archivematica popular among various digital preservation organizations. Collaborators in the development of Archivematica include the City of

---

[101] Artefactual, "Archivematica Homepage," https://www.archivematica.org/en/, last accessed September 28, 2016; Van Garderen, et al., "The Archivematica Project," pp. 5-6.
[102] Archivematica Wiki, "Micro-Services," https://wiki.archivematica.org/Micro-services, last accessed September 28, 2016.
[103] Artefactual, "Preservation Planning," https://www.archivematica.org/en/docs/archivematica-1.4/user-manual/preservation/preservation-planning/#preservation-planning, last accessed September 28, 2016.
[104] Artefactual, "Ingest," https://www.archivematica.org/en/docs/archivematica-1.4/user-manual/ingest/ingest/#ingest, last accessed September 28, 2016.

Vancouver Archives,[105] Museum of Modern Art, University of Alberta Libraries, Rockefeller Archive Center, and many others.[106]

Preservica

Preservica is an OAIS compliant cloud-based digital preservation program created by Preservica Digital Preservation, a company that is part of the Tessella group. Tessella is a software services, information analytics, and technology consulting company that specializes in solving complex technological problems.[107] In April 2014, Tessella rebranded its digital preservation service (previously known as Safety Deposit Box) as Preservica and created Preservica Digital Preservation as a "separate, wholly owned subsidiary company backed by the Tessella group."[108] Much like Archivematica, Preservica contains OAIS compliant workflows and the ability to automate ingest.[109] The program also allows for easy customization of the user interface. This allows users to search for records within Preservica and also use the database across multiple devices including tablets and smartphones.[110] There is also the ability to add extensive amounts of metadata to digital records as well as robust and customizable storage and security features.[111]

---

[105] The City of Vancouver Archives was one of the first institutions to partner with Artefactual in the development phases of Archivematica, see Van Garderen, et al., "The Archivematica Project," p. 3.
[106] Archivematica Wiki, "Homepage," https://wiki.archivematica.org/Main_Page, last accessed September 28, 2016.
[107] Tessella, "Tessella Launches Preservica Subsidiary Business," http://web.archive.org/web/20140517013140/http://tessella.com/press-releases/tessella-launches-preservica-subsidiary-business, last accessed October 28, 2016.
[108] Tessella, "Tessella Launches Preservica Subsidiary Business," http://web.archive.org/web/20140517013140/http://tessella.com/press-releases/tessella-launches-preservica-subsidiary-business, last accessed October 28, 2016.
[109] Preservica, "How Preservica Works," http://preservica.com/preservica-works/, last accessed September 28, 2016.
[110] Preservica, "Public access and discovery," http://preservica.com/public-access-discovery/, last accessed September 28, 2016.
[111] Preservica, "How Preservica Works," http://preservica.com/preservica-works/, last accessed September 28, 2016.

Currently, there are three editions of Preservica. They are the Cloud, Standard, and Enterprise editions. The Cloud edition is ideal for smaller organizations that do not have access to extensive IT resources because the data is stored by the Preservica provider. It should be noted that while this version of Preservica is useful for organizations that do not have the IT resources to develop their own storage, digital records may be subject to additional legal and administrative challenges. Standard edition is ideal for medium sized organizations that have access to IT resources and is stored and managed by the user. Finally, there is Enterprise Edition which is for large organizations with extensive IT resources that can manage and fully customize the Preservica software to suit their needs. All three of these editions contain fully OAIS compliant workflows.[112]

The decision to use the cloud has led to many challenges with implementing Preservica. Organizations that store their digital records within Preservica's cloud may have their records stored in a foreign country. Preservica has two cloud servers, one in the eastern United States and the other in Dublin, Ireland. Depending on where the records are stored, they will be subject to local laws and regulations instead of, or in addition to, those of the country of origin. Examples can include having records subject US laws such as the PATRIOT Act and access laws. Organizations that do not want their records to be subject to laws of other nations may be hesitant to store their records in the cloud.[113] In addition, Preservica's cloud also contains the records of many organizations and while work is done to minimize risks, there is always a possibility that records from one organization may accidently be available to other

---

[112] Preservica, "Preservica Editions," http://preservica.com/editions-pricing/.

[113] Kevin O'Farrelly, Alan Gairey, James Carr, Maïté Braud, Robert Sharpe, and Ann Keen, "Access and Preservation in the Cloud: Lessons from Operating Preservica Cloud Edition," http://purl.pt/26107/1/DLM2014_PDF/19%20-%20Access%20and%20Preservation%20in%20the%20cloud%20.pdf, p.2, last accessed September 23, 2016.

organizations.[114] Finally, being able to remove records from the cloud is important as an organization may wish to change service providers. Preservica addresses this issue by allowing users to download all their records and metadata from the cloud in order to transfer it to another provider.[115]

Despite the challenges associated with the cloud based functionality of Preservica, there is a heavy emphasis on ease of use and customizability. This has no doubt helped in its widespread adoption by many organizations. One such organization is the Wellcome Library in London, England. Preservica was used to help preserve medical documents for the UK Medical Heritage Library (UK-MHL) Project. The goal of the UK-MHL Project was to provide free access to digitised copies of records relating to medical history. By using many systems in addition to Preservica, the Wellcome Library was able to preserve many made digital records.[116] During this project, other systems were used in addition to Preservica. For example, the Wellcome Library used a system called Goobi to harvest data from the Internet Archive, do metadata mapping, and create METS files. When a digital record is ingested into Preservica, the administrative metadata that is created is then transferred and merged into Goobi's database. This metadata is stored by the institution for future use.[117] This implementation, however, showed that Preservica has some issues. In her experience using Preservica at the Wellcome Library, Victoria Sloyan describe the difficulties of removing digital records from Preservica. When a digital record is ingested into Preservica, a catalogue record is created. The problem

---

[114] O'Farrelly et al., "Access and Preservation," p. 2.
[115] O'Farrelly et al., "Access and Preservation," p. 2.
[116] Christy Henshaw, Dave Thompson, João Baleia, "Automating Harvest and Ingest of the Medical Heritage Library," https://www-ariadne-ac-uk.uml.idm.oclc.org/issue73/henshaw-et-al/, (2015), last accessed September 28, 2016.
[117] Henshaw, et al., Automating Harvest."

occurs when ingested files fail appraisal. This requires staff to manually remove records from Preservica, which is a long and tedious process.[118]

Preservica has also been deployed by a variety of organizations including the National Archives in Britain, Alabama Department of Archives and History, Museum of Modern Art, and several others.[119] Reasons for adopting Preservica as the primary digital preservation system are varied. One reason was that storage services could not be created in-house which made Preservica's cloud based approach attractive.[120] Another reason for adopting Preservica was the ability to add accession information and descriptive data to digital records.[121] Cloud storage seems to be a significant determiner in using Preservica. For an organization that may not have access to extensive IT resources to create their own cloud, this can be beneficial. The downside to this, of course, is that the archives loses some of its control over its own records. If the cloud servers are located across international borders, the archives' records will be subject to foreign laws. Furthermore, archives with sensitive information may not prefer to have their records stored overseas and in a public cloud where there is a small risk that other organizations may accidently have access to them.

The technical and legal difficulties associated with cloud computing aside, how well does Preservica address the challenges outlined in Chapter One? To recap, digital preservation is challenging because digital records rely on hardware and software to become usable. If the hardware, software, or both degrades or becomes obsolete, the digital record may become

---

[118] Victoria Sloyan, "Born-digital Archives at the Wellcome Library: Appraisal and Sensitivity Review of Two Hard Drives," *Archives and Records*, 37:1 (2016): pp. 24-25.
[119] Patricia Franks, "Government Use of Cloud-Based Long Term Digital Preservation as a Service: An Exploratory Study," *IEEE 2015 Digital Heritage International Congress* (Sept-Oct 2015): p. 3; Jon Tilbury, "Making Digital Preservation Part of the Digital Preservation Lifecycle," *PASIG Conference, San Diego* (2015): http://web.stanford.edu/group/dlss/pasig/PASIG_March2015/20150313_Presentations/Tilbury_Preservica.pdf, last accessed September 28, 2016.
[120] Franks, "Government Use," p. 3.
[121] Franks, "Government Use," p. 3.

inaccessible. These challenges are in addition to the challenges of selecting records worth preserving and preserving valuable contextual information. Through its use of cloud computing, Preservica is able to store digital records that are easily available. A significant drawback to this, of course, is that an archive's digital records may be stored in another country and is managed by a third party. For archives that contain sensitive information, IT resources would be needed to create in-house servers. Preservica also offers a robust digital preservation service but it is important to keep in mind that it is just another digital preservation tool. The selection of digital records that are worth preserving and the kinds of metadata to include in preservation are choices made by archivists. Preservica does not make digital preservation policy, archivists do. To conclude, while Preservica is a robust digital preservation system, it is not a replacement for careful planning and clear digital preservation policies.

OAIS Compliance Within Archivematica and Preservica

Both Archivematica and Preservica heavily promote the fact that they are compliant with digital preservation standards such as OAIS, but what does this mean? What exactly makes something "OAIS compliant?" The definition of OAIS compliance is not clear. In Schumann and Recker's article on the myth of OAIS compliance, the authors describe how complying with OAIS means complying with standards that are abstract and open to interpretation.[122] As we have seen when we examined OAIS, the standard only provides a basic framework in which to operate. OAIS does not provide specific and concrete examples of how to implement its recommendations. There are no clear instructions on how to create information packages or how to work with a designated community that is not to have a consistent knowledge base. All of this

---

[122] Schumann and Recker, "De-Mystifying OAIS," p. 6.

is up to the user's own interpretation of how to set up an OAIS and modify the standard to suit an organization's needs. Due to the amount of interpretation required to make OAIS workable, Schumann and Recker argue that OAIS compliance should be measured in degrees rather than a simple compliant or non-compliant measurement as it is too problematic to determine how well someone adheres to standards that are already up to interpretation.[123] The implementation of OAIS within Archivematica and Preservica are their respective developer's interpretations of OAIS.

Artefactual felt that OAIS compliance was important because it was seen as "the de-facto standard for designing digital archives systems."[124] Therefore, OAIS became the template for building Archivematica. Archivematica uses OAIS workflows such as ingest and access as well as concepts such as AIPs, SIPs, and DIPs.[125] Through using Archivematica, I have worked through these different parts of OAIS' implementation. SIPs are ingested, AIPs are generated with metadata, and a DIP is created for access purposes. This is done through the use of microservices. Archivematica's version of OAIS is simple and functional. Since Archivematica is free to download and try, it acts as a nice showcase of what an OAIS compliant system can accomplish. I describe the use of Archivematica in greater detail in Chapter Three.

Preservica takes a similar approach to Archivematica in terms of OAIS compliance. Preservica has built a system that is compliant with OAIS. Like Archivematica, Preservica uses OAIS terms and workflows. SIPs are ingested, AIPs are created and preserved, and DIPs are generated for access.[126] Other OAIS functionalities that are present in Preservica are data management, administration, preservation planning, and access. It is worth noting that unlike

[123] Schumann and Recker, "De-Mystifying OAIS," p. 10.
[124] Van Garderen et al., "The Archivematica Project," p. 4.
[125] Van Garderen et al., "The Archivematica Project," pp. 5-6.
[126] O'Farrelly et al., "Access and Preservation," pp. 3-4.

Archivematica, Preservica is not free to download. Therefore, it cannot act as a free demonstration to those who are interested in seeing an OAIS compliant system.

Archivematica and Preservica share similarities when it comes to their definitions of OAIS compliance. Both digital preservation systems use information packages and the workflows described in OAIS. They differ in how they implement this compliance. Archivematica does this through microservices while Preservica accomplishes it through a single system that is hosted on the cloud. Archivematica is also more easily accessible than Preservica as it is free to download and has an active user community. Preservica is a closed proprietary program that is run solely by Tessella. This openness makes Archivematica a much more accessible showcase for an OAIS compliant system. It should be noted that this OAIS compliance is Artefactual's or Tessella's interpretation of it. Another company may have a slightly different view on the matter as OAIS served as a guideline rather than a roadmap for a specific implementation.

PREMIS Compliance within Archivematica and Preservica

As we have seen with OAIS, PREMIS is often abstract and does not provide a clear step-by-step implementation guide. Organizations often have to modify PREMIS and/or their organizations,[127] in addition to using other metadata standards, to implement PREMIS.[128] Just like OAIS compliance, PREMIS compliance is different depending on how one interprets the PREMIS data dictionary. Both Archivematica and Preservica promote the fact that they allow for the input of PREMIS metadata. Interestingly, little has been written on how Archivematica and Preservica use PREMIS within their digital preservation systems. Archivematica allows

[127] Donaldson and Conway, "Implementing PREMIS," pp. 283-286.
[128] Wilson, " How Much is Enough," pp. 213-214.

users to input PREMIS metadata during the transfer and ingest stage of the preservation workflow.[129] In my experience using Archivematica, the preservation metadata is accessible by accessing the AIP and opening the XML file. The metadata describes aspects of the record such as what microservices were used and what format it was normalized in. It also includes metadata elements that describe the different events, agents, rights, and restrictions associated with the digital record.[130] This preservation metadata is created automatically when the SIP is ingested into Archivematica. While large amounts of preservation metadata are generated, it is not possible to modify the XML file if it is necessary to migrate to a new format. Furthermore, PREMIS metadata within Archivematica is not actionable. This means that while Archivematica can automatically generate PREMIS compliant metadata, it cannot make rules within the system. For example, I cannot program access restrictions based around PREMIS' rights metadata. All that can be currently done with Archivematica is to generate the metadata. Despite this issue, Archivematica provides organizations with a free tool for automatically generating vast amounts of PREMIS compliant preservation metadata. This is important as it freely highlights the impact that preservation metadata has on a record's provenance. As we have seen in Chapter One, digital preservation activities often change the record and documenting this change is required for maintaining a record's authenticity and integrity. PREMIS provides archives with a tool to do this and this is why it is important within Archivematica. Preservica also uses PREMIS metadata in its packaging and metadata schema called XIP. In addition, Preservica takes into account the different entities within PREMIS.[131] As with OAIS, there will be variations in how

---

[129] Artefactual, "PREMIS Metadata in Archivematica," https://www.archivematica.org/en/docs/archivematica-1.4/user-manual/metadata/premis/, last accessed September 28, 2016.
[130] A listing of PREMIS metadata elements used in Archivematica can be found here: https://wiki.archivematica.org/Metadata_elements, last accessed September 28, 2016.
[131] Mark Evans, "XIP and PREMIS," https://www.loc.gov/standards/premis/pif/2012/PREMIS%20Fair%202012-EVANS.pdf, last accessed September 28, 2016.

PREMIS is used by organizations and developers of digital preservation systems due to PREMIS being implementation independent. I should also note that the proprietary nature of Preservica means that I am unable to assess how metadata is used within the system.

Importance of Compliance

As we have seen, compliance with OAIS and PREMIS is highly promoted in the field. A quick browse through Archivematica's and Preservica's websites will reveal how much these products advertise their compliance to these standards despite the fact that their compliance is based on their own interpretations of these documents. Indeed, why is it so important to be compliant with standards that are so open to interpretation? What OAIS and PREMIS have accomplished is to provide archivists with a framework and common vocabulary with which to work. This common understanding of digital preservation is why OAIS and PREMIS have been so widely adopted by the digital preservation community. Working within the same framework and using the same terms is why compliance with these standards is so important and provides a standard with which to measure. It will allow for much more effective collaboration and communication between organizations as they will all be working from the same document. How the concepts of PREMIS and OAIS are applied is specific to each institution, but this common framework is why compliance with these two standards is significant.

Conclusion

Many methods have been developed to remedy the challenges of digital preservation. Migration and data redundancy have been created to deal with obsolescence and degradation while influential standards such as OAIS and PREMIS have been developed to provide guidance to institutions who are creating a digital preservation program. Preservation systems such as Preservica and Archivematica have also been developed to facilitate digital preservation. All of these standards and systems, however, are far from perfect as they do not fix the root cause of digital preservation challenges: the reliance on hardware and software to render digital records. Digital records will always rely on software and hardware that will eventually become obsolete and inaccessible due to the constant advancement of technology. As a result, information professionals will constantly need to make sure that older digital records are accessible with new technology.

How we go about doing this, however, is open to interpretation. Standards such as OAIS and PREMIS provide guidance and programs such as Archivematica and Preservica can provide the necessary tools, but how they are used is up to the individual organization. There is no clear path that an organization can follow when developing their digital preservation program other than their own. An organization knows better than anyone what its needs are and how best meet them. The tools and standards are just means of achieving an institution's goal of preserving digital records. This can often lead to different repositories having slightly different digital preservation systems, as we will see in the next chapter, where I will explore digital preservation methods in more depth by conducting several case studies of how these tools are used.

Chapter Three

Archivematica Case Study at the Mennonite Heritage Centre Archives

As we have seen in the previous chapter, there have been many standards and products developed to address the challenges of digital preservation. While standards such as OAIS and PREMIS provide guidance for developing digital preservation programs, their applications are often up to the interpretation of the user. Systems such as Archivematica and Preservica are also helpful tools for managing and preserving digital records. Using digital preservation standards such as PREMIS and OAIS, they can identify the file format of a digital object and migrate it to a preservation format. These programs also create large amounts of metadata on digital objects that are valuable for preservation efforts. How does this work in the real world? In a time when sufficient amounts of funding are not guaranteed, how do organizations successfully preserve their records?

The difficulty of securing a steady stream of funding is a major challenge in digital preservation. Kastellec has identified the lack of stable funding sources as a significant difficulty for digital preservation. The costs of maintaining the technological capacity of properly managing digital records can be difficult to sustain over long periods of time. Paying for the time of even a limited number of staff is also a significant challenge faced by organizations.[1] In the article "Archival Digital Preservation Programs: Staffing, Costs, and Policy," Shelby Sanett describes the financial difficulties of maintaining a digital preservation program and the difficulties of staffing projects. Sanett found that several organizations had difficulty in

[1] Kastellec, "Practical Limits," pp. 67-68.

determining the actual costs of digital preservation.[2] This challenge is faced by all organizations and is amplified by the increasing volume of digital records that require digital preservation.[3]

Therefore, in this chapter I personally experiment with various preservation methods to see which ones best preserve digital records on a minimal budget. I conducted a case study at the Mennonite Heritage Centre Archives (MHCA) where I ran an implementation of Archivematica and tested how well it functioned with its digital records. I also installed Archivematica on my personal computer and tested it with my own digital records. I chose to run Archivematica over other digital preservation systems because it is free for anyone to download and use. This is important to anyone who wants to conduct digital preservation but cannot afford expensive tools or does not have access to extensive IT resources. I will also examine other case studies to see what other organizations have done to address the challenge of digital preservation. This chapter examines what has been done and what archives and other organizations can do to preserve digital records.

Archivematica at the MHCA and at Home

The Mennonite Heritage Centre Archives (MHCA) was founded in 1929 as an inter-Mennonite facility for the preservation of records relating to Mennonite history. The MHCA also serves as the primary repository for various Mennonite organizations including the Mennonite Church Canada, Canadian Mennonite Board of Colonization, the Evangelical Mennonite Mission Conference, and many others.[4] The MHCA also specializes in the

---

[2] Shelby Sanett, "Archival Digital Preservation Programs: Staffing, Costs, and Policy," *Preservation, Digital Technology & Culture* 42 (2013): pp. 144-146.
[3] Sanett, "Archival Digital Preservation Programs," pp. 138-139.
[4] Mennonite Heritage Centre Archives, "About Mennonite Heritage Centre Archives," http://archives.mennonitechurch.ca/About, last accessed September 28, 2016; Mennonite Heritage Centre Archives, "History of the Mennonite Centre Archives," http://archives.mennonitechurch.ca/history, last accessed September 28, 2016.

preservation and management of various historical documents regarding Prussian and Russian Mennonite communities as well as various congregational, individual, and family records relating to the Mennonite community.[5] As the primary archive for so many institutions and communities, the MHCA has a wide variety of records in different media including textual, photographic, audio, and many others. Much like other institutions, the MHCA is expecting an increase in the amount of digital records that it accepts into its custody. As a result, it needs to develop a plan for managing the influx of digital records and so were willing to give me permission to test Archivematica.

Archivematica was chosen due to its free cost. My focus during this case study is the records that were donated for the project by the Mennonite Church Canada's Communication Department. The Communication Department's primary responsibility is to promote Mennonite Church Canada and to liaise with the public. Various digital records were contributed to the project by the department including textual, photographic, video, and audio files in a variety of formats. This provided an excellent test sample for Archivematica. Prior to the beginning of the case study, the MHCA purchased a new computer to run Archivematica and test its functionalities. The version of Archivematica that I used for this test study was version 1.0. This was the first full release of Archivematica. The case study was conducted in April and May 2014.

During the course of this case study, Archivematica's performance was mixed. First and foremost, Archivematica's user interface is intuitive and easy to use. It is a simple task to select digital records for ingest and to move them throughout the preservation workflow. The program also easily creates large amounts of descriptive and preservation metadata for the ingested

[5] Mennonite Heritage Centre Archives, "About Mennonite Heritage Centre Archives," http://archives.mennonitechurch.ca/About, last accessed September 28, 2016.

records with little effort on the part of the user. This minimal user interaction required in Archivematica makes digital preservation a relatively easy experience but it did have its flaws. While textual and most photographic and audio files could be ingested quite easily, video files were highly problematic. As part of the Communication Department's functions, it took high quality video of many events hosted by Mennonite Church Canada. Some of the longer files can be many gigabytes in size. Through many tests, Archivematica could not process information packages that were over three gigabytes without crashing. Archivematica still crashed in some tests where there was only one file that was over three gigabytes.

While these challenges were problematic, the biggest difficulty that I encountered with Archivematica was installing it. Out of seven installation attempts, only one was successful. This lone success lasted for about a month before it too crashed from running out of space on the virtual hard drive. Before installing Archivematica, several additional programs need to be downloaded and installed. First, VirtualBox needs to be installed.[6] This is a program that runs a virtual machine that allows for additional operating systems to run on a computer without overwriting the machine's existing operating system. Next, the Ubuntu operating system[7] needs to be installed in VirtualBox because Archivematica is not programmed to work with the Windows operating system. Instead, it was only programmed to work with an open source Linux operating system like Ubuntu. Once these programs have been installed, the Archivematica program and the packages required by Archivematica need to be installed within the Ubuntu virtual machine. These packages include database software such as MySQL and search software like ElasticSearch. This process of installing Archivematica within the Ubuntu virtual machine proved to be highly problematic. I had several problems such as the MySQL

[6] VirtualBox, https://www.virtualbox.org/, last accessed September 28, 2016.
[7] Ubuntu, http://www.ubuntu.com/, last accessed September 28, 2016.

server crashing and ElasticSearch failing to install. These failures made Archivematica unworkable. There is an Archivematica Google Group that provides assistance with troubleshooting technical difficulties,[8] but I found that having the technical experience required for solving problems within Ubuntu to be an essential asset for dealing with problems. Unless an organization is willing to pay for Artefactual's service, having in-house IT support is highly desirable.

During the course of the case study, I decided to experiment with an additional digital preservation program since I could not successfully install Archivematica. The program I decided to test on the Communication Department's records was XENA. XENA was developed by the National Archives of Australia and it is free to use and install. XENA works by scanning digital objects and normalizing them into XENA files which require the XENA Viewer to access. A major advantage that I found to using XENA is that it is simple to install. All that is required is to download XENA and follow the prompts from the installer. It also works with Windows and does not require the creation of a virtual machine. XENA was also not prone to crashing when handling larger files. Being able to install it onto Windows, however, leads to a problem. If a new release of Windows is incompatible with the XENA Viewer, all XENA files will be inaccessible. As a result, there is a constant threat of obsolescence. Despite its ease of use, XENA is also nowhere near as feature-rich as Archivematica. Unlike Archivematica, XENA does not create vast quantities of descriptive and preservation metadata when it processes digital records. This can prove problematic for anyone wanting to do more than just create XENA files. Indeed, XENA is also not OAIS and PREMIS compliant. There are no information packages or preservation metadata elements. XENA's sole function is to normalize the digital

[8] Archivematica Google Group, https://groups.google.com/forum/#!forum/archivematica, last accessed September 28, 2016.

object into a XENA file. This does not address the digital preservation challenges that were raised in Chapter One. If XENA were ever to become obsolete, there would be no other way to open the XENA files, which makes XENA not a viable digital preservation system. While XENA is much more stable than Archivematica, it is also not as robust.

In the November 2015, I decided to test the latest version of Archivematica (version 1.4) on my home computer. My computer had similar specifications to the one used at the MHCA. I began by downloading VirtualBox and Ubuntu.[9] I then created an Ubuntu virtual machine within VirtualBox. Once Ubuntu was working, I started installing Archivematica.[10]

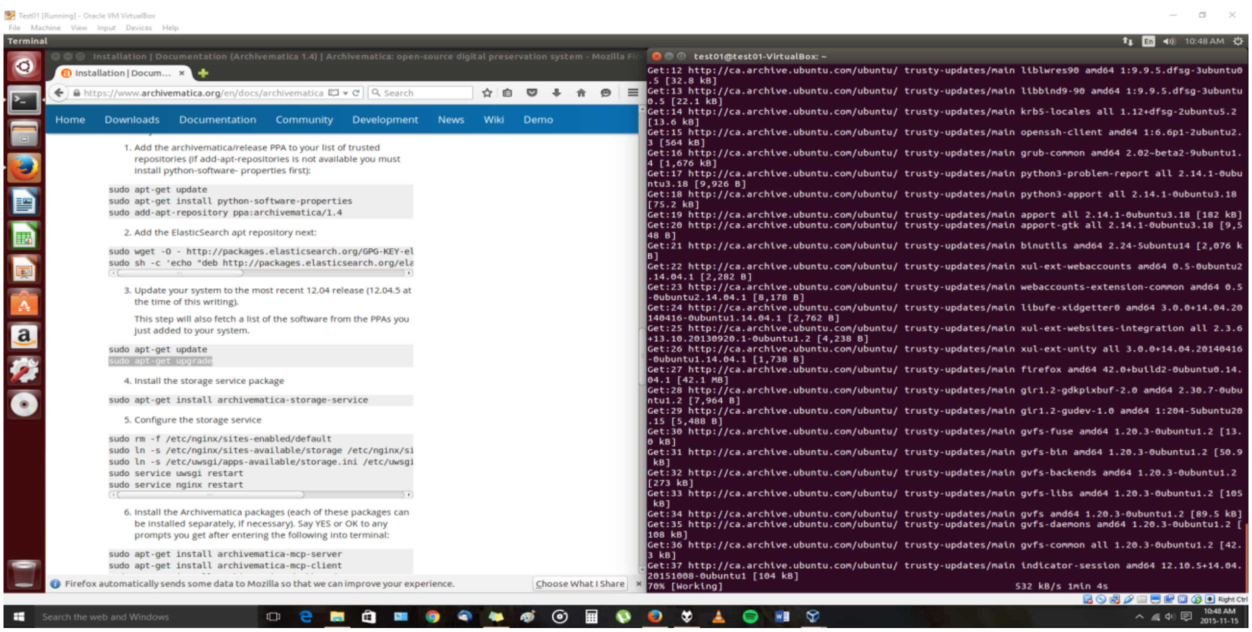

Figure 1: Installing Archivematica

[9] VirtualBox, https://www.virtualbox.org/wiki/Downloads, last accessed September 28, 2016; Ubuntu, http://www.ubuntu.com/desktop, last accessed September 28, 2016.
[10] Downloading instructions for Archivematica can be found here: https://www.archivematica.org/en/docs/archivematica-1.4/admin-manual/installation/installation/, last accessed September 28, 2016.

As we can see in Figure 1, installing Archivematica requires downloading the Archivematica packages through the Terminal application in Ubuntu. While this process may look intimidating to individuals who may not be technically experienced, it only involves minimal input from the user. The user need only copy and paste the commands from Archivematica's website (shown on the left in Figure 1) to the Terminal. The user then selects yes or okay to any prompts that come up during installation.

Near the end of the installation process, the user needs to start the Archivematica Storage Service and Dashboard. The Storage Service manages the locations and workflows within Archivematica while the Dashboard performs the preservation tasks. To start the Dashboard, the user need to enter the requested information as shown in Figure 2.

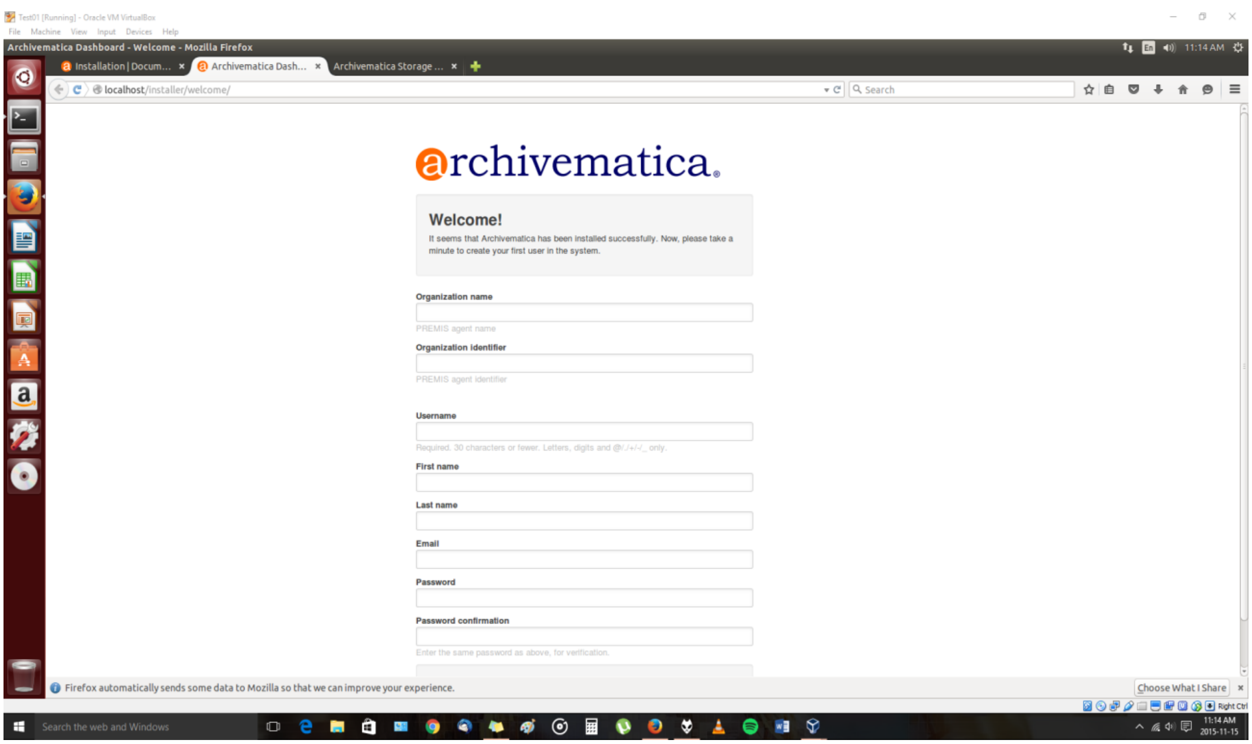

Figure 2: Starting the Dashboard

Once the information has been input into Archivematica, users will have access to the Archivematica Dashboard. As shown in Figure 3, the Dashboard is minimalistic and easy to use.

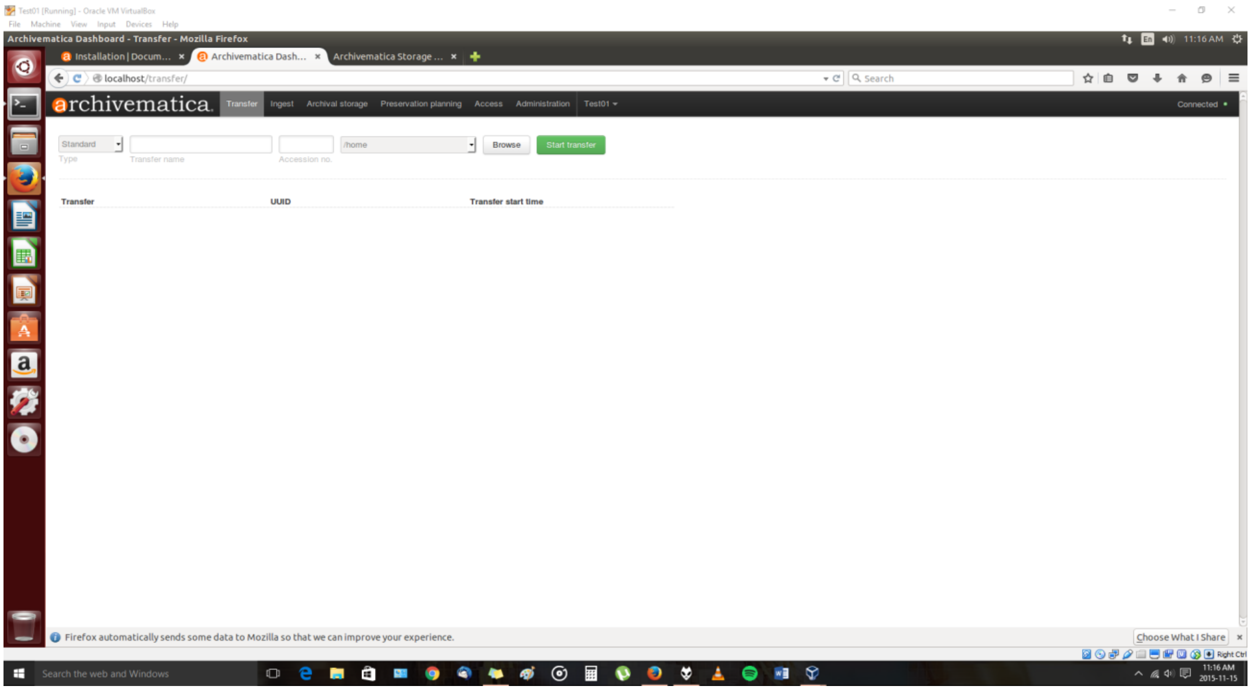

Figure 3: Archivematica Dashboard

Ingesting collections is also relatively easy in Archivematica. By selecting Browse and navigating to the folder that holds the required files, users can easily begin the ingest process.

As seen in Figure 4, there is little human interaction required.

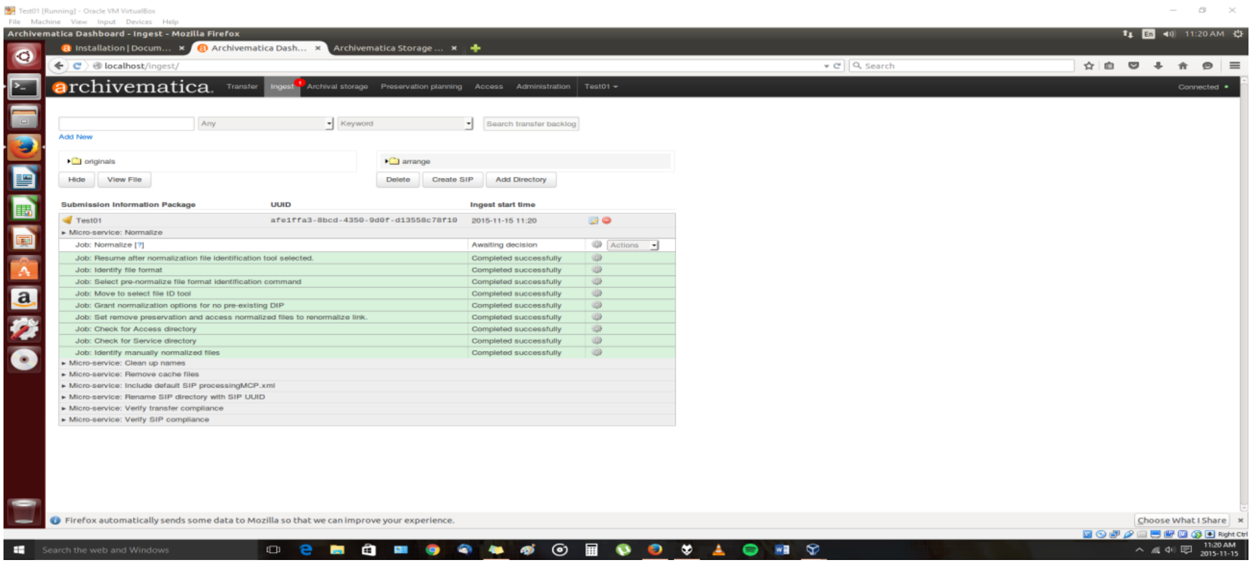

Figure 4: Ingest Process

As seen in Figure 5, users also have the option to input descriptive metadata into their transfers.

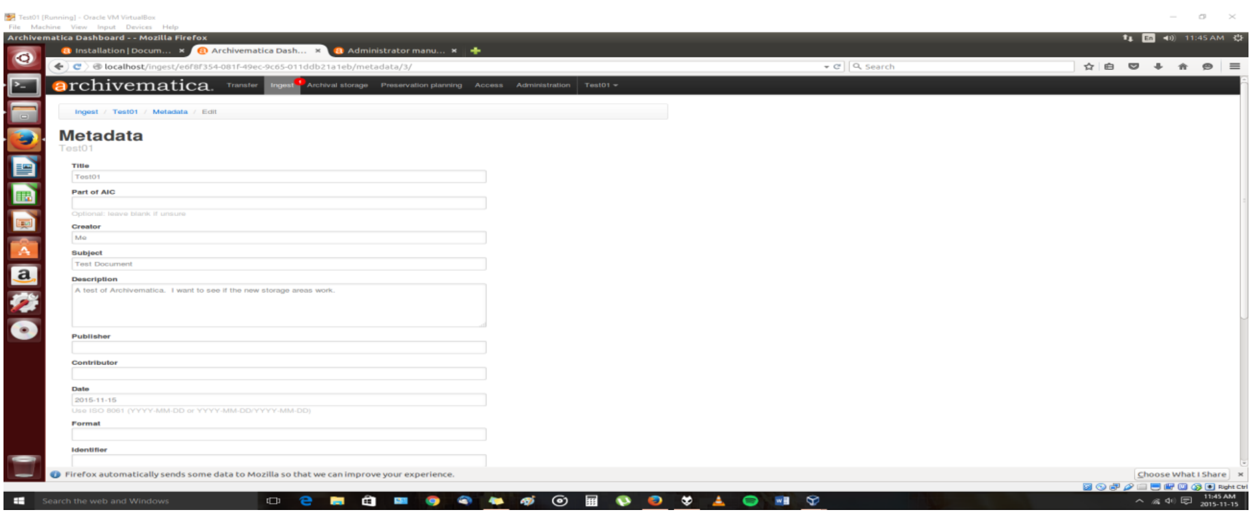

Figure 5: Metadata

After ingest is complete, the items are moved to the transfer phase. As seen in Figure 5, this phase is similar to the ingest process in that it too requires little action on the part of the user.

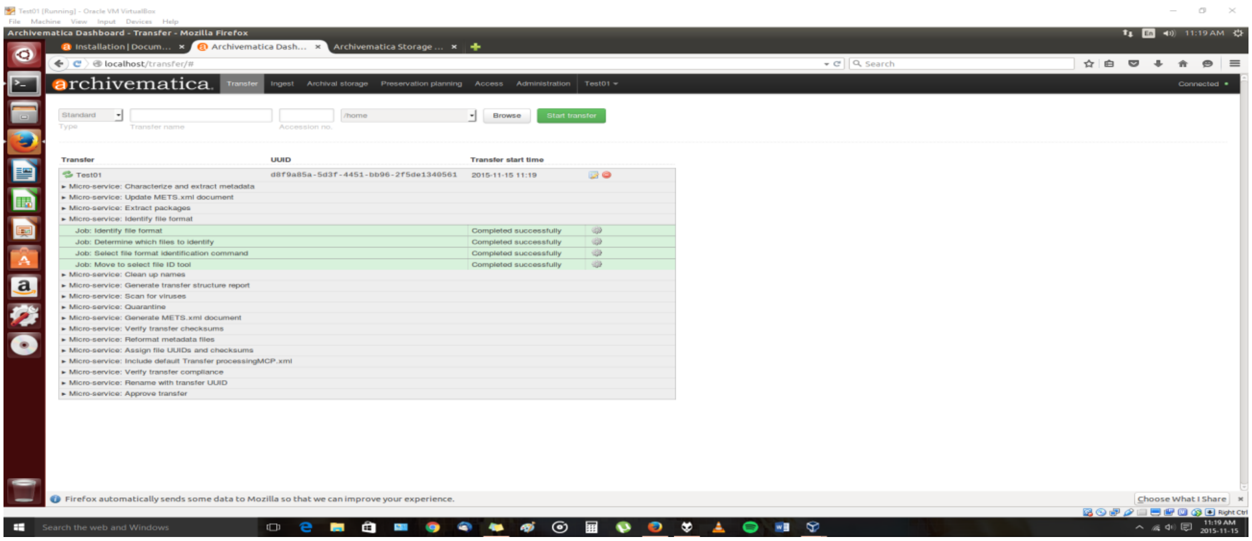

Figure 6: Transfer in Archivematica

Once the transfer process is complete, the items have been successfully processed by Archivematica.

While there were no problems with this installation, I did find that there were many problems when modifying Archivematica. To access the Archival Information Packages (AIPs) and Dissemination Information Packages (DIPs) without having to go through the Archivematica Dashboard and Storage Service, the user must go into the Archivematica directory within the var folder. This folder is locked and cannot be accessed regardless of the user's administration privileges. Changing the storage location is easy to do in the Storage Service, but this leads to many problems within Archivematica's workflows. The most significant one is that AIPs and DIPs could not be created as the Dashboard did not recognize the storage location as valid. After trying to change to workflows that match the new storage destination, I still could not get Archivematica to work properly. I ended up having to reinstall it.

Another problem is that Elasticsearch would crash when I tried to start the Dashboard. This required me to restart Elasticsearch by using the Terminal application. The most significant problem, however, is that I had an “Internal server error” after I created shared folders between my virtual and host machines whenever I tried to start the Archivematica Dashboard.

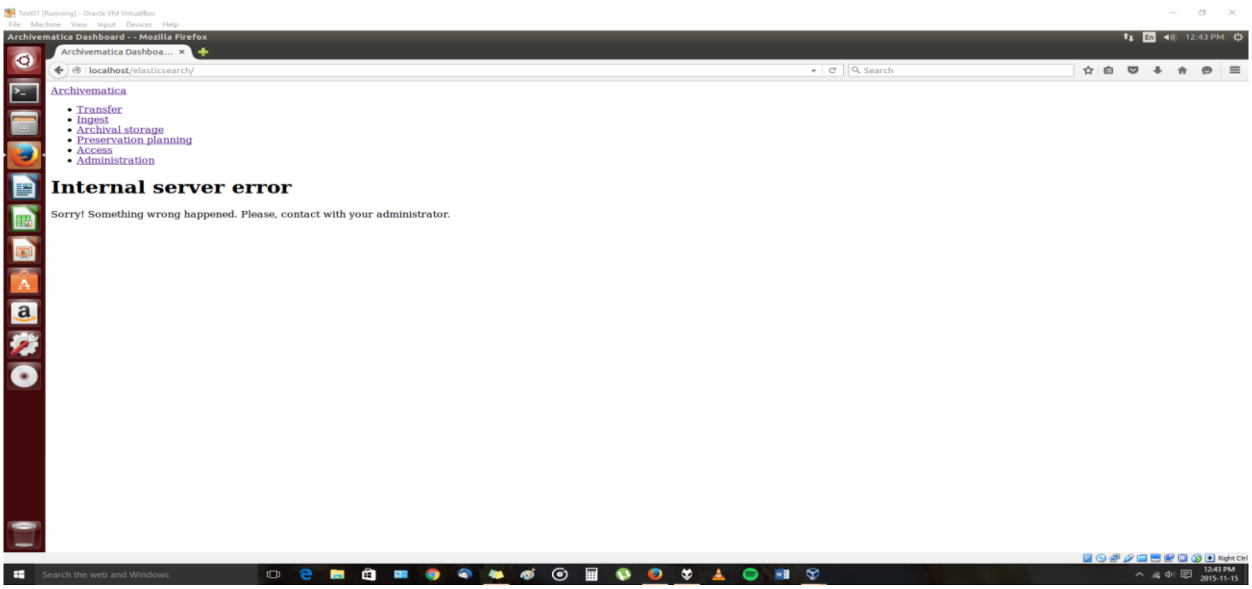

Figure 7: Internal server error

A shared folder is required when using Archivematica because without it, it will be impossible to transfer files between the virtual and host machines because they are, effectively separate computers. A good way to think about this is that I am running a second computer within my computer. This leads to the problem of not being able to access the files on my host computer when I am in the virtual machine and vice versa. Shared folders need to be created using VirtualBox to transfer files between the host and virtual machines. After trying all the fixes that I know, I could not find a solution for this issue. As a result, this installation of Archivematica became unusable.

I decided to do a second installation of Archivematica on a new virtual machine. After installing Ubuntu, I created the shared folder and then installed Archivematica. The installation was successful but got the same internal server error whenever I tried to change the Transfer Source in the Storage Service to access the content on the shared folder. I tried restarting the MCP server and client as well as the Gearman Job Server to no effect. As a result, this installation of Archivematica also became unusable.

Through my work at the MHCA and on my home computer, I have learned many important lessons about running a digital preservation program on a minimal budget. First, IT resources or at least access to IT support is a major asset for a digital preservation program. These resources can be used to address any of the technical problems that arise with glitches and bugs in these programs. Secondly, the more powerful a computer that an archive is are able to acquire, the better these programs will run. The MHCA acquired a brand new computer that had impressive hardware specifications, but I was only able to process an accession in Archivematica that was under three gigabytes in size. If I had a less powerful machine at my disposal, this number may have been much lower. Finally, the most important lesson that I learned during this case study is that Archivematica or any digital preservation program is not a replacement for a digital preservation and information management policy. Digital preservation programs are tools that are meant to be used within a wider policy for digital records. They do not replace the need for managing digital records and selecting ones that are worth preserving and the ones that can be disposed of when their retention has expired. As we have seen in Chapter One, managing the vast quantities of digital records that are created daily is a huge challenge that digital preservation systems such as Archivematica do not address. These must be solved by the individuals involved in creating and administering digital preservation programs. In short, digital preservation systems such as Archivematica are just tools to be used within an organization's digital preservation program. This is, of course, assuming that an archives can get them to function properly.

Archivematica, the Challenges of Digital Preservation, and Standards

Throughout my work with Archivematica at the MHCA and at home, I considered how well it handles the challenges that I describe in Chapter One and how it satisfies the standards in Chapter Two. As we have seen in Chapter Two, Archivematica promotes the fact that it is OAIS compliant and that it generates PREMIS compliant metadata. To recap the challenges described in Chapter One, digital preservation is difficult due to digital records' reliance on hardware and software. When either the software, hardware, or both become obsolete or degrades, the digital record will become unusable. This challenge is in addition to selecting what is worth preserving out of the vast amounts of digital records that are created daily. Finally, there is the challenge of preserving context through the use of various types of metadata.

On its own, Archivematica does not satisfy all of these challenges as it just a single tool in an organization's digital preservation program. It is not a replacement for sound planning and consistent policies. Therefore, this analysis of Archivematica will assume that it is being used as part of a wider digital preservation strategy. In this scenario, Archivematica is used at an organization that has developed a digital preservation policy. This digital preservation policy outlines which records will be selected and which file formats the archive will accept. It will also be assumed that the organization in this scenario will have minimal access to IT resources as this is quite common at small to medium sized archives.[11] Finally, I will note that this section is heavily informed by my experience using Archivematica at home and at the MHCA.

Before any digital records can be accepted by the archives, it must be determined which records will be taken in by the archives. As seen in Chapter One, large quantities of digital records are created daily and in a variety of file formats. Determining which ones are worthy of preservation and which ones are not is one of the human challenges in digital preservation. A few things need to be considered when appraising digital records including how many files to take, their organization, and what formats they are in. I agree with Mumma et al. that appraisal of digital records should be done at a practical level by considering what archivists can afford to keep.[12] Determining what an archive can afford to keep, however, is reliant on many factors. First among them is whether the archive can preserve the types of digital records being processed. All digital archives should know what formats they are capable of preserving as this will affect the types of digital records acquired. In the case of Archivematica, it can process a variety of file formats. Its Format Policy Registry outlines the formats Archivematica can process, what formats the digital object are changed to, and the tools used during normalization.[13] The wide range of formats that Archivematica can process is one of its strengths. The Archivematica Format Policy can serve as a guide for determining what file formats to accept.

A weakness that Archivematica has, however, is in storage. As seen in my experience at the MHCA, the reason my only successful install of Archivematica failed is because the virtual

---

[11] Due to lack of funding and technical expertise, several cultural organizations lack significant IT resources to assist with digital preservation. See Tom Evens and Laurence Hauttekeete, "Challenges of Digital Preservation for Cultural Heritage Institutions," *Journal of Librarianship and Information Science* 43:3 (2011).
[12] Mumma et al., "A First Look," p. 110.
[13] Artefactual, "Format Policies," https://wiki.archivematica.org/Format_policies, last accessed September 28, 2016.

hard drive reached maximum capacity. When I tried to enlarge it, Archivematica kept crashing, making that install unusable. This problem was in addition to the fact that I was only able to process no more than three gigabytes of data at one time. In Chapter One, we saw that digital records are vulnerable to hardware degradation and should be copied to multiple storage devices through the process of data redundancy. In my experience with Archivematica, this is not an easy task to accomplish. When I tried to change the storage location in the Storage Service, this caused many errors and made the program unusable. Thus, additional copies of digital records cannot be easily created using Archivematica. This is a significant flaw from a digital preservation standpoint because if the computer that Archivematica is running on malfunctions, all data could be lost. Therefore, the IT resources of the archives need to be considered as they can assist with solving technical issues such as these.

A second consideration when appraising digital records is whether the archives has the permissions from the copyright holder to change records. As we have seen in Chapter One, many digital preservation actions involve making copies and changes to digital records in order to make them readable on current machines. This may violate copyright laws if the archive does not have these rights from the copyright owner of the digital records. Archivematica creates additional copies and normalizes digital objects. These actions may break copyright law. To avoid any legal hurdles to digital preservation, the archives must acquire ownership of the copyright or permissions from the copyright owner of the records that it is processing. This can be done through clear donor agreements that transfer ownership to the archives or provide the rights to modify the record for preservation purposes. Furthermore, these issues do not just apply to contents of digital records, they also apply to the software that is required to render them. The software will become obsolete and in order to modify them for preservation purposes, permissions will have to be sought from the copyright holder. An example would be modifying a program to access digital objects. Permissions from whomever owns the software's copyright would have to be granted before any modifications could take place. In my experience at the MHCA, the records created by the Mennonite Church Canada's Communication Department were the property of Mennonite Church Canada. As the MHCA was the main repository of the Mennonite Church Canada, the MHCA had full rights to perform whatever preservation actions are necessary. These legal considerations are important to keep in mind as they can easily prevent any digital preservation actions from taking place.

When the digital records have been accepted by the archives, Archivematica can begin processing. As we have seen in the images above, Archivematica has an intuitive user interface that is easy to navigate. It is also easy to select digital objects for preservation and initiate workflows. All of this is done with minimal user interaction. Indeed, the only time the user is needed is to select different options and input descriptive metadata when prompted. The ability to add descriptive and preservation metadata is a significant feature when compared to other freely available digital preservation programs such as Xena, which has no such feature. As with other aspects of Archivematica's user interface, adding descriptive metadata is easy as it only involves filling out the required fields. As seen in figure 5 above, adding descriptive metadata is not unlike regular archival description. This makes adding metadata simple even for archivists who may not be particularly technically experienced.

These easy to use features are available for free download from Artefactual's website. This makes Archivematica an attractive option for archivists, but the system does have significant flaws when processing large digital objects. Smaller digital objects like emails, spreadsheets, and word processing documents are easily processed by Archivematica. It is larger documents such as video and high definition images that Archivematica has difficulty processing. During my time at the MHCA, I could not process any transfer that was over three gigabytes without Archivematica crashing. This was especially problematic when I was processing high definition video files that were many gigabytes in size. With additional IT resources, an archives may be able to solve this problem, but it remains a difficult challenge nonetheless.

How well does Archivematica solve the challenges discussed in Chapter One? While it is intuitive, has excellent metadata features, can process many file formats, and is freely available, Archivematica has many drawbacks that reduce its effectiveness in combatting digital preservation challenges. First, while Archivematica generates PREMIS compliant metadata, this data is not actionable. This means that I cannot develop rules within Archivematica that are based on the metadata that it had created. The metadata also cannot be modified and the only thing that can be done is to read it. Archivematica also cannot process large digital objects without crashing. This makes processing any video files very problematic. Furthermore, it is challenging to change the storage options in the Storage Service without leading to many error messages when processing digital objects. This leads to difficulties in making additional copies to prevent data loss if disaster were to strike the archive. While these problems could be corrected with additional IT support, any organization that does not have access to that kind of support will have significant difficulties in implementing Archivematica. In short, Archivematica is held back from completely addressing the challenges described in Chapter One by its own technical difficulties.

While it is held back by its own technical difficulties, how well does Archivematica comply with OAIS and PREMIS? As we have seen in Chapter Two, OAIS compliance is a nebulous concept. OAIS itself is a abstract standard that is be open to interpretation. When Artefactual promotes Archivematica's OAIS compliance, it is really promoting what it considers to be OAIS compliance. What Artefactual considers to be OAIS compliance may not be what others consider OAIS compliance. As for PREMIS, Archivematica allows users to generate PREMIS compliant preservation metadata. It should be noted that like OAIS, how PREMIS is implemented is open to interpretation and the same can be said about the nebulous nature of PREMIS compliance. Archivematica has PREMIS compliant metadata based on what Artefactual considers to be PREMIS compliance. Another organization may have different ideas about what constitutes PREMIS compliant metadata. Indeed, this highlights the imprecise nature of compliance with PREMIS.

Conclusion

This study of Archivematica's utility in handling the challenges of digital preservations highlights the importance of having a clear and consistent digital preservation policy. Decisions need to be made on what digital records should be preserved, how they should be preserved, and how they should be stored. All of these tasks should be outlined in the archive's digital preservation policy. In addition, decisions on how much descriptive and preservation metadata is needed should also be outlined in this policy. Archivematica does not address any of these concerns. It is only a digital preservation tool and is not a replacement for a digital preservation policy.

While Archivematica is free and intuitive to use, it is held back by its technical limitations and requires significant IT support to make it a feasible digital preservation program. It is important to note, however, that Archivematica itself is a digital record that is vulnerable to the same digital preservation challenges outlined in Chapter One. It too is vulnerable to degradation and obsolescence.

Much like other systems, Archivematica is dependent on specific software and hardware to operate. Archivematica runs on the Ubuntu operating system, which is regularly updated when new versions are released. As a result, Archivematica may not be able to run on newer versions of the operating system. To get around this issue, Artefactual has instructions on how to upgrade an installation of Archivematica,[14] but what happens when a new virtual machine needs to be created because the current version of Archivematica becomes incompatible with a version of Ubuntu? Since I encountered difficulties in changing storage locations on Archivematica, I assume that it will be difficult to get data out of Archivematica. Admittedly, I am not a software engineer and do not posses the technical expertise to answer that question, but this issue should be considered when deciding to use Archivematica.

To conclude, Archivematica is a robust and easy to use digital preservation system that is held back by technical difficulties. If an archive does not have access to significant IT resources, Archivematica will be a challenging program to operate. Even with IT resources, Archivematica should be used as part of a wider digital preservation plan.

[14] Artefactual, "Installation," https://www.archivematica.org/en/docs/archivematica-1.4/admin-manual/installation/installation/#install-1-4, last accessed September 28, 2016.

Conclusion

Digital records are important means of communication. While the creation and use of analog records will continue, the use of digital records has become important in how we interact with each other. As a result, there are many digital records that are created daily and this excessive volume,[1] as well as the reliance of hardware and software, has made preserving digital records challenging. This thesis provided an overview of these challenges and the responses to them. I explained that since digital records have specific software and hardware requirements, they are vulnerable to hardware degradation and obsolescence. The volume of digital records also makes it difficult to decipher what is valuable from what has little to no value. Once something of value has been found, it is even more challenging to ensure that it remains meaningful with various types of metadata and that the essence of what gives the records value is not lost in migration.

Digital preservation techniques such as migration and emulation have been developed to make obsolete digital records accessible. Format migration changes the obsolete file format of a digital record to one that is currently used and emulation involves recreating a digital object's original operating environment. Standards such as OAIS and PREMIS as well as digital preservation systems such as Archivematica and PREMIS have been developed to assist with mitigating the preservation challenges posed by digital records. But how effective are they? Do they ensure that digital records will be accessible for as long as they are required?

There is no current method that will "solve" the problems associated with preserving digital records. Technology will always continue to advance and this will inevitably lead to

[1] To provide an illustration of how many digital records are created daily, see Steven Lewis' website "One Second on the Internet," (http://onesecond.designly.com/, last accessed September 28, 2016). It provides a helpful visual of how many digital records are created every second on the Internet. This graphic does not take into account the many digital records that are not online.

older technologies becoming obsolete. As a result, digital records that rely on these outdated technologies will be vulnerable to obsolescence. No matter how many times a digital record is migrated to a current format, this format will eventually become obsolete and a new format will be needed. With each new migration, there is always a chance that valuable content or significant properties may be lost in the transition. Through several migrations, a record can become something different from what it originally was. While emulation does avoid migration, the emulation itself is vulnerable to obsolescence as it too is a digital record that relies on specific hardware and software to function. As a result, new emulations need to be built when old ones become obsolete. All of this is assuming that someone can determine which digital records are considered worthy of preservation out of the vast quantities that are created daily. In short, there is no permanent solution at the moment to the challenges of digital preservation.

This seems like a rather bleak picture of digital preservation but, while no available solution solves all of these problems, they can be managed. Even though standards such as OAIS and PREMIS are far from perfect because of their abstract wording and lack of concrete implementation plans, they do provide some valuable guidance for organizations seeking to preserve their digital records. As seen in chapter two, these standards are used by many institutions to communicate with each other using a common vocabulary when discussing digital records. This communication and networking between institutions is valuable because, as we have seen in chapter one, no single institution has the resources (both financial and human) to archive all digital records. Partnerships should be made to allow institutions to focus resources on preserving what matters to them.

Focus, foresight, and planning are the keys for managing digital records. An organization should focus on preserving records that are considered to have long-term value and records that have little value should only be preserved for as long as necessary. To accomplish this goal, archivists should modify their appraisal policies to not only select digital records worthy of preservation, but also what can realistically be preserved by the archive. A preservation strategy will outline the steps needed to preserve digital records including what types of metadata is needed. Following these steps will allow the archive to plan preservation actions and reduce the risks of losing digital records. As we have seen in chapters two and three, digital preservation systems such as Archivematica and Preservica are valuable tools for any digital preservation program but they are not replacements for careful planning.

While having carefully crafted digital preservation and appraisal plans will not provide a permanent solution to the challenges of digital preservation, it can assist with preserving them for as long as they are needed. These plans involve carefully selecting digital records for preservation and then determining their significant properties. Once this is done, backing up the digital records, adding administrative, preservation, and descriptive metadata, and determining the appropriate digital preservation strategy based on the significant properties are needed. There is, however, no one size fits all plan to implement such a strategy. As we have seen in chapter two, even OAIS and PREMIS do not give concrete examples of how to go about doing this. An organization's financial and human resources also need to be taken into account. As demonstrated in my case study in Chapter Three and my own experiences using Archivematica, an archive that does not have access to IT resources will have a difficult time running a large-scale digital preservation project. All institutions, however, can at least do something to handle the challenges of digital preservation. Ignoring the problem is the worst thing that an archivist at any institution can do.

An archivist can develop a strategy to select digital records for preservation through appraisal. All institutions will have different needs and requirements for this, but the point is to focus on preserving the digital records that are important and that can be preserved. A valuable way to accomplish this is to create a records management program that clearly outlines the retention periods and file formats of digital records.[2] With such a plan, the archivist can focus on preserving specific digital records for the long term while only preserving others until their retention period expired. As we have seen in chapter one, developing format policies that outline what file formats digital records can be saved in assists in digital preservation by limiting the amount of file formats that an archive has to preserve. With proper training in digital preservation archivists can create digital preservation programs that fit their institutions' needs and resources. These policies and procedures will also assist in managing digital records and facilitate finding information. They will also remove many digital records that have little value, thereby making digital preservation a much easier task.

Therefore, archivists need to be proactive in their approach to digital records and move away from the traditional "custodial view" of archives. Archivists should be involved in the writing of records management and file policies as this will help ensure that records of enduring value are saved in formats that are widely supported. There should also ideally be frequent audits of digital records to ensure their authenticity during their retention periods. A "pre-custodial"[3] approach to digital preservation and record management should be adopted as this

[2] Admittedly, this will be most useful for organizational records rather than personal archives. Individuals, however, could modify some of these concepts to better suit their personal digital preservation needs. An example of this is format management. Saving files in formats that are less likely to become obsolete in the near future is beneficial for digital preservation. Likewise, saving records on storage media that is up to date and unlikely to quickly become obsolete is another beneficial strategy. Steps such as these will greatly assist with the preservation of personal digital records.

[3] Please see footnote 38 of the Introduction for the definition of the pre-custodial approach.

will allow archives to preserve and assist with the management of records prior to ingest into the archives. This is beneficial because some records have long retention periods that may require preservation actions to ensure that the record remains accessible for as long as necessary.

An active archives will also raise the visibility of the repository within the wider organizational structure. The archives becomes an integral part of the digital records management program because they conduct preservation activities and ensure that records remain accessible for as long as necessary. This increased involvement also means more partnership between the archives and different departments or organizations such as information technology to assist with technical matters where the archives may lack the resources to handle them by itself. This greater organizational role also will have the result of showing the importance of the archives as they will be required to ensure that digital records are not lost. The hopeful result from this greater presence is increased funding for archives, which in turn will allow the hiring of additional staff to manage the complex digital preservation issues of which there are many.

While any institution can backup its digital records to removable hard drives, larger amounts of data will require greater technical experience. Creating and managing vast quantities of administrative, description, and preservation metadata is also a challenge as it is unrealistic to think that staff have the time to manually add metadata to digital records. Thus, the ability to automatically generate preservation and descriptive metadata is a neccessity. Indeed, this is one of the strengths of Archivematica that we have seen in chapters two and three. However, while digital preservation systems such as Archivematica can be valuable assets for digital preservation, they also need sufficient technical expertise to install and operate. Migrating digital records and creating/running emulations can also prove to be difficult to implement if the technical infrastructure does not exist within an organization. Finally, there is the challenge of ensuring that significant properties are not lost throughout migrations.

Some of these challenges can be mitigated by an under-resourced institution through careful planning. With records management and file format policies, archivists can greatly increase the amount of time needed before migration by selecting for preservation digital records that are in formats that are widely used and supported. Since these formats are so common, they will likely be readable on most computers for sometime. Even when they are no longer supported, as we have seen in chapter one, an open source renderer may be made to view these file formats. This, however, is an imperfect solution as there is no guarantee that an open source renderer will be developed or for how long the renderer will be supported by current machines. As we have seen with emulation, software is also a digital record. Additional training in the use of these digital preservation systems and methods is needed.

To conclude, there is no current solution to the challenges of digital preservation. This is due to digital records relying on hardware and software to be readable by humans. With a few exceptions, most analog records do not require additional technologies to be used by researchers. Since digital records rely on technology, every time technology advances there is the probablility of it becoming obsolete and not being able to run on current file formats. Nothing short of halting all technological advancement will stop this reoccurring risk. In many ways, the challenges of digital preservation are similar to living with an illness that cannot be cured. With medication, the symptoms of the illness can be managed but not entirely cured. In digital preservation, technology will continue advancing and leaving behind obsolete digital records that cannot run on current machines. Through the intervention of archivists, digital records can remain accessible with the use of digital preservation practices such as migration and systems such as Archivematica. They will not cure the problems of digital preservation, but they will manage it and make records that would have otherwise been lost available to all users.

In short, the current means of digital preservation do save records that would have been lost to obsolescence or degradation, but they do not solve the underlying problems that plague digital records. How do we preserve a record in a format or storage medium that will become obsolete? It may take a decade or two, or it could happen in only a few years, but that record will eventually become obsolete. At the moment, this cycle of obsolescence and change is not solved by our preservation solutions. What is available at the moment, however, allows archivists at the very least to ensure that the record is not lost. It can be accessible, albeit in a format that may be different from the original. Archivists can manage digital records so that they can be accessible for as long as required. This, after all, is the ultimate goal of preservation.

Bibliography

Allison, Arthur, James Currall, Michael Moss, and Susan Stuart. "Digital Identity Matters," *Journal of the American Society for Information Science and Technology* 56:4 (February 15, 2005): pp. 364–372.

Anderson, David. "Preserving Europe's Digital Cultural Heritage: A Legal Perspective." *New Review of Information Networking* 18:1 (May 2013): pp. 16–39.

Ankerson, Megan Sapnar. "Writing Web Histories with an Eye on the Analog Past." *New Media and Society* 14 (2012): pp. 384-400.

Archivematica, https://www.archivematica.org/en/, last accessed September 28, 2016.

Auditor General of Canada. "Chapter 7 Documentary Heritage of the Government of Canada – Library and Archives Canada." Fall 2014.

Bak, Greg. "Trusted By Whom? TDRs, Standards Culture, and the Matter of Trust." *Archival Science* 16:6 (Fall 2016).

Bass, Jordan. "A PIM Perspective: Leveraging Personal Information Management Research in the Archiving of Personal Digital Records." *Archivaria* 75 (Spring 2013): pp. 49-76.

Bass, Jordan. "Getting Personal: Confronting the Challenges of Archiving Personal Records in the Digital Age." Master's thesis, University of Manitoba, (2012).

Bearman, David. "Moments of Risk: Identifying Threats to Electronic Records." *Archivaria* 62 (2006): pp. 15-46.

Bearman, David. "Record-Keeping Systems." *Archivaria* 36 (1993): pp. 16-36.

Bell, Alan. "Standards and Standards Culture: Understanding the Nature and Criticisms of Standardisation," in *Comma* (Paris, 2001).

Bennion, Jeff. "The Rage-inducing Word versus WordPerfect Debate," http://abovethelaw.com/2014/06/the-rage-inducing-word-versus-wordperfect-debate/, last accessed September 28, 2016.

Besk, June M., Jessica Coates, Brian Fitzgerald, Wilma Mossink, William G. LeFurgy, Adrienne Muir, Mary Rasenberger, Christopher D. Weston. "Digital Preservation and Copyright: An International Study." *The International Journal of Digital Curation* 3:2 (2008): pp. 103-111.

Breytenbach, Amelia and Ria Groenewald. "The Use of Metadata and Preservation Methods for Continuous Access to Digital Data." *The Electronic Library* 29:2 (2011): pp. 236-248.

Brown, Adrian. *Practical Digital Preservation: A How-to Guide for Organizations of Any Size* (London, 2013).

Campbell-Kelly, Martin. "The History of the History of Software." *IEEE Annals of the History of Computing* (2007): pp. 40-51.

Caplan, Priscilla. "Understanding PREMIS," http://www.loc.gov/standards/premis/understanding-premis.pdf, (February 2009), last accessed September 28, 2016.

Caplan, Priscilla and Rebecca Guenther. "Practical Preservation: The PREMIS Experience." *Library Trends* 54:1 (Summer 2005): pp. 111-124.

Carroll, Laura, Erika Farr, Peter Hornsby, and Ben Ranker. "A Comprehensive Approach to Born-Digital Archives." *Archivaria* 72 (Fall 2011), pp. 61-92.

Cerf, Vinton G. "Avoiding "Bit Rot": Long-Term Preservation of Digital Information." *Proceedings of the IEEE* 99:6 (June 2011), pp. 915-916.

Cohen, Frederick B. "Digital Diplomatics and Forensics." *Records and Information Management Journal*, 25:1 (March 2015): pp. 21-44.

Comeau, Martin. "Archives, Historical Climate Records, and the Climate Observations of Thomas Corcoran, Hudson's Bay Company, 1827-1841." Master's thesis, University of Manitoba, (2005).

The Consultative Committee for Space Data Systems (CCSDS). "About CCSDS." http://public.ccsds.org/about/default.aspx, last accessed September 28, 2016.

CCSDS. *Reference Model for an Open Archival Information System (OAIS)*. June 2012.

Cook, Terry. "Easy to Byte, Harder to Chew: The Second Generation of Electronic Records Archives." *Archivaria* 33 (Winter 1991-92): pp. 202-216.

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

Coyle, Karen. *Rights in the PREMIS Data Model*. http://www.loc.gov/standards/premis/Rights-in-the-PREMIS-Data-Model.pdf, last accessed September 28, 2016.

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

Cunningham, Adrian. "Waiting for the Ghost Train: Strategies for Managing Electronic Personal Records Before its too Late." *Archival Issues* 24:1 (1999): pp. 55-64.

Cunningham, Adrian. "Digital Curation/Digital Archiving: A View from the National Archives of Australia." *The American Archivist* 71:2 (2008): pp. 530-543.

Dappert, Angela and Adam Farquhar. "Significance Is in the Eye of the Stakeholder." *Research and Advanced Technology for Digital Libraries* eds. Maristella Agosti, José Borbinha, Sarantos Kapidakis, Christos Papatheodorou, Giannis Tsakonas (Berlin 2009).

Dappert, Angela, Sébastien Peyard, Carol C.H. Chou, and Janet Delve. "Describing and Preserving Digital Object Environments." *New Review of Information Networking* 18:2 (2013): pp. 106-173.

Donaldson, Devan Rey and Paul Conway. "Implementing PREMIS: A Case Study of the Florida Digital Archive." *Library Hi Tech* 28:2 (2010): pp. 273-289.

Donaldson, Devan Ray and Elizabeth Yakel. "Secondary Adoption of Technology Standards: The Case of PREMIS." *Archival Science* 13:1 (May 11, 2012): pp. 55-83.

Duranti, Luciana. "Reliability and Authenticity: The Concepts and their Implications." *Archivaria* 39 (Winter 1994-95), pp. 5-10.

Duranti, Luciana and Heather MacNeil. "The Protection of the Integrity of Electronic Records: An Overview of the UBC-MAS Research Project." *Archivaria* 42 (Fall 2006): pp. 46-67.

Elves, Daniel. "Advocating Electronic Records: Archival and Records Management Promotion of New Approaches to Long-Term Digital Preservation." Master's thesis, University of Manitoba, (2012).

Evans, Mark. "XIP and PREMIS." https://www.loc.gov/standards/premis/pif/2012/PREMIS%20Fair%202012-EVANS.pdf, last accessed September 28, 2016.

Evens, Tom and Laurence Hauttekeete. "Challenges of Digital Preservation for Cultural Heritage Institutions." *Journal of Librarianship and Information Science* 43:3 (2011).

Franks, Patricia. "Government Use of Cloud-Based Long Term Digital Preservation as a Service: An Exploratory Study." *IEEE 2015 Digital Heritage International Congress* (Sept-Oct 2015): pp. 371-374.

Galloway, Patricia. "Personal computers, microhistory, and shared authority: Documenting the inventor-early adopter dialectic." *IEEE Annals of the History of Computing* 33:2 (2011), pp. 60-61.

Garret, John and Donald Waters. *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information*. Council on Library and Information Resources (Washington D.C., 1996), http://www.clir.org/pubs/reports/pub63watersgarrett.pdf, last accessed September 28, 2016.

Giaretta, David. *Advanced Digital Preservation*, Springer (Berlin, 2011).

Geist, Michael, http://www.michaelgeist.ca/tag/digital-locks/, last accessed September 28, 2016.

Gilliand, Anne J. "Archival Appraisal: Practicing on Shifting Sands" in *Archives and Recordkeeping: Theory into Practice*, ed. Caroline Brown, (London 2014): pp. 31-62.

Guttenbrunner, Mark, Christoph Becker and Andreas Raubner. "Keeping the game alive: evaluating strategies for the preservation of console video games." *International Journal of Digital Curation* 5:1 (2010), pp. 64-90.

Harrison, Donald Fisher. "Computers, Electronic Data, and the Vietnam War." *Archivaria* 26 (Summer 1988): pp. 18-32.

Harvey, Ross. *Preserving Digital Materials*. De Gruyter (2012).

Harvey, Ross and Martha Mahard. "Mapping the Preservation Landscape for the Twenty-First Century." *Preservation, Digital Technology & Culture*, 42:1 (2013): pp. 5-16.

Hedstrom, Margaret. "Digital Preservation: A Time Bomb for Digital Libraries." *Computers and Humanities* 31:3 (1997), pp. 189-202.

Henshaw, Christy, Dave Thompson, and João Baleia, "Automating Harvest and Ingest of the Medical Heritage Library," https://www-ariadne-ac-uk.uml.idm.oclc.org/issue73/henshaw-et-al/, (2015), last accessed September 28, 2016.

InterPARES. "The Preservation of the Integrity of Electronic Records." http://www.interpares.org/UBCProject/intro.htm, last accessed September 28, 2016.

Kastellec, Mike. "Practical Limits to the Scope of Digital Preservation." *Information Technology and Libraries* 31:2 (2012), pp. 63-71.

Krahn, Konrad. "Looking Under the Hood: Unravelling the Content, Structure, and Context of Functional Requirements for Electronic Record Keeping Systems." Master's thesis, University of Manitoba, (2012).

Laughton, Paul. "OAIS Functional Model Conformance Test: A Proposed Measurement." *Program: Electronic Library and Information Systems* 46:3 (2012), pp. 308-320.

Lavoie, Brian F. *The Open Archival Information System Reference Model: Introductory Guide*. (January 2004).

Lee, Christopher A., Matthew Kirschenbaum, Alexandra Chassanoff, Porter Olsen, and Kam Woods. "BitCurator: Tools and Techniques for Digital Forensics in Collecting Institutions." http://www.dlib.org/dlib/may12/lee/05lee.html.

Lee, Christopher A. "Open Archival Information System (OAIS) Reference Model." *Encyclopedia of Library and Information Studies, Third Edition* (2010): pp. 4020-4030.

Levi, Charles. "Five Hundred 5.25-inch Discs and One (finicky) Machine: A Report on a Legacy e-records Pilot Project at the Archives of Ontario." *Archivaria* 72 (2011), pp. 239-246.

Library and Archive Canada. *Local Digital Format Registry*.

Library of Congress. *PREMIS Version 3.0*. http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf, last accessed September 28, 2016.

Library of Congress. *Recommended Format Specifications*. http://www.loc.gov/preservation/resources/rfs/TOC.html, last accessed September 28, 2016.

Lillico, Kristina. "Preparing for the Digital World in the Federal Government." *Navigating the Digital Future IM Conference*, (November 27, 2013): http://www.im.gov.ab.ca/documents/conference/2.1_Preparing_for_the_Digital_World_-_KLillico.pdf, last accessed September 28, 2016.

Mahoney, Michael. "The Histories of Computing(s)." *Interdisciplinary Science Reviews* 30:2 (2005), pp. 119-135.

Mahoney, Michael. "What Makes the History of Software Hard." *IEEE Annals of the History of Computing* (July-September 2008), pp. 8-18.

Marsden, Paul. "When is the Future? Comparative Notes on the Electronic Record-Keeping Projects of the University of Pittsburgh and the University of British Columbia." *Archivaria* 43 (Spring 1997), pp. 158-173.

McDonough, Jerome P. "'Knee-Deep in the Data:' Practical Problems in Applying the OAIS Reference Model to the Preservation of Computer Games." *45th Hawaii International Conference on System Science* (2012), pp. 1625-1634.

McKemmish, Sue. "Evidence of Me." *The Australian Library Journal* 45:3, pp. 174-187.

Mumma, Courtney C., Glenn Dingwall, and Sue Bigelow. "A First Look at the Acquisition and Appraisal of the 2010 Olympic and Paralympic Winter Games Fonds: or, SELECT * FROM VANOC_Records As Archives WHERE Value="true"." *Archivaria* 72 (Fall 2011): pp. 93-122.

O'Donnell, Kye. "Taming Digital Records with the Warrior Princess: Developing a Xena Preservation Interface with TRIM." *Archives and Manuscripts*, 38:2 (November 2010), pp. 37-60.

O'Farrelly, Kevin, Alan Gairey, James Carr, Måt̊é Braud, Robert Sharpe, and Ann Keen. "Access and Preservation in the Cloud: Lessons from Operating Preservica Cloud Edition."

http://purl.pt/26107/1/DLM2014_PDF/19%20-%20Access%20and%20Preservation%20in%20the%20cloud%20.pdf, last accessed September 28, 2016.

Online Computer Library Center (OCLC) and Research Libraries Group (RLG). *A Metadata Framework to Support the Preservation of Digital Objects*. OCLC Online Computer Library Center (Dublin, Ohio: 2002): http://www.oclc.org/content/dam/research/activities/pmwg/pm_framework.pdf, last accessed September 28, 2016.

Perrin, Joy, Heidi Winkler, Le Yang. "Digital Preservation Challenges with an ETD Collection – A Case Study at Texas Tech University." *The Journal of Academic Librarianship* 41 (2015): pp. 98-104.

Preservica, http://preservica.com/, last accessed September 28, 2016.

Ramesh, P., J. Vivekavardhan, and K. Bharathi. "Metadata Diversity, Interoperability and Resource Discovery Issues and Challenges." *DESIDOC Journal of Library & Information Technology* 35:3 (April 22, 2015): pp. 193-199.

Rinehart, Amanda Kay, Patrice-Andre Prud'homme, and Andrew Reid Huot. "Overwhelmed to Action: Digital Preservation Challenges at the under-Resourced Institution." *OCLC Systems & Services* 30:1 (February 2014): pp. 28-42.

Rosenthal, David S.H. "Bit Preservation: A Solved Problem?" *The International Journal of Digital Curation* 1:5 (2010), pp. 134-148.

Rosenthal, David S.H. "Format Obsolescence: Assessing the Threat and the Defenses." *Library Hi Tech* 28:2 (2010), pp. 195-210.

Rosenthal, David S.H. "Keeping Bits Safe: How Hard Can it Be?" *Communications of the ACM* 53:11 (November 2010), pp. 47-55.

Rosenzweig, Roy. "Scarcity or Abundance? Preserving the Past in a Digital Era." *American Historical Review* 108:3 (June 2003): pp. 736-762.

Rothenberg, Jeff. "Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation, A Report to the Council on Library and Information Resources." (January 1999), pp. 1-35.

Sanett, Shelby. "Archival Digital Preservation Programs: Staffing, Costs, and Policy." *Preservation, Digital Technology & Culture* 42 (2013): pp. 137-149.

Schumann, Natascha and Astrid Recker. "De-Mystifying OAIS Compliance: Benefits and Challenges of Mapping the OAIS Reference Model to the GESIS Data Archive." *IASSIST Quarterly* 36:2 (2012): pp. 6-11.

Sheldon, Madeline. *Analysis of Current Digital Preservation Policies: Archives, Libraries, and Museums*, (2013). http://www.digitalpreservation.gov/documents/Analysis%20of%20Current%20Digital%20Preservation%20Policies.pdf, last accessed September 28, 2016.

Sloyan, Victoria. "Born-digital Archives at the Wellcome Library: Appraisal and Sensitivity Review of Two Hard Drives." *Archives and Records*, 37:1 (2016): pp. 20-36.

Snowdon, David A. "Healthy Aging and Dementia: Findings from the Nun Study." *Annals of Internal Medicine* 139, (2003): pp. 450-454.

Spence, Jaqueline. "Preserving the Cultural Heritage: An investigation into the Feasibility of the OAIS Model for Application in Small Organisations." *Aslib Proceedings: New Information Perspectives* 58:6 (2006), pp. 513-524.

Tilbury, Jon. "Making Digital Preservation Part of the Digital Preservation Lifecycle." *PASIG Conference, San Diego* (2015): http://web.stanford.edu/group/dlss/pasig/PASIG_March2015/20150313_Presentations/Tilbury_Preservica.pdf, last accessed September 28, 2016.

Thibodeau, Kenneth. "Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years," in *The State of Digital Preservation: An International Perspective* (July 2002), pp. 4-31.

United Nations Educational, Scientific and Cultural Organization. "Guidelines for the Preservation of Digital Heritage," (March 2003).

Van Garderen, Peter, P. Jordan, T. Hooten, C. Mumma, and E. McLellan. "The Archivematica Project." http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/VC_Van_Garderen_et_al_26_Workshop1.pdf, last accessed September 28, 2016.

Vardigan, Mary and Cole Whiteman. "ICPSR Meets OAIS: Applying the OAIS Reference Model to the Social Science Archive Context." *Archival Science* 7:1 (2007): pp. 73-87.

Victorian Electronic Records Strategy. http://prov.vic.gov.au/government/vers, last accessed September 28, 2016.

Vielfaure, Natalie. "Medical Records Redefined: The Value of the Archival Record in Medical Research." Master's thesis, University of Manitoba, (2015).

Wilson, Andrew. "How Much is Enough: Metadata for Preserving Digital Data." *Journal of Library Metadata*," 10 (2010): pp. 205-217.

Yeo, Geoffrey. "'Nothing Is the Same as Something Else': Significant Properties and Notions of Identity and Originality." *Archival Science* 10:2 (June 10, 2010): pp. 85-116.

Zweig, Ronald. "Electronically Generated Records and Twentieth Century History." *Computers and The Humanities* 27 (1993): pp. 73-83.