

Practical Digital Solutions:
A DIY Approach to Digital Preservation

by
Tyler McNally

A Thesis Submitted to the Faculty of Graduate Studies of
the University of Manitoba
In Partial Fulfillment of the Requirements of the Degree of

MASTER OF ARTS

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

Copyright © 2018 by Tyler McNally

Table of Contents

Abstract.....	i
Acknowledgments.....	ii
Acronym Index.....	iii
Introduction.....	1
Chapter 1.....	12
Chapter 2.....	54
Chapter 3.....	81
Conclusion	116
Bibliography.....	125

Abstract

Since the introduction of computers, archivists have had to find ways to deal with digital records. As more records are born digital (created through digital means) and digital technologies become more entrenched in how data is created and processed, it is imperative that archivists properly preserve these records. This thesis seeks to propose one possible solution to this issue. Rather than advocate for paid solutions or electronic record management systems, it advocates for more practical in-house DIY solutions. The first chapter lays out background information and the historiography of digital archiving in Canada at the federal level. The second chapter moves step-by-step through a workflow developed at the University of Manitoba's Faculty of Medicine Archives that lays out one possible DIY style solution. The third chapter is an audit of the workflow from the second chapter against three important international standards for preserving digital information.

Acknowledgments

I would like to acknowledge and thank Professors Thomas Nesmith and Greg Bak. Their role as professors of the Archival Studies program has been a great source of support and inspiration as well as their knowledge and passion for both archives and their students. I would also like to especially thank Greg for his role as my thesis advisor and not only helping me to write and edit this thesis, but for Greg's own work with digital archiving and computer history that has served as a source of inspiration for my own thoughts and ideas about digital archiving.

I would also like to thank Jordan Bass for his role as a mentor. Without him this thesis would not have been possible as it was under his guidance and direction that I was able to put together the original workflow. His role as a mentor figure to myself has been key in developing how I think and approach digital archiving.

Further I would like to thank my friends and family that have supported me throughout this process. Without them letting me vent, bounce ideas off them and their encouragement this thesis would not have been possible

Acronym List

AIP	Archival Information Package
CCSDS	Consultative Committee for Space Data Systems
COPPUL	Council of Prairie and Pacific University Libraries
CRL	Center for Research Libraries
DIP	Dissemination Information Package
IMOSA	Information Management and Office Systems Advancement
ISO	International Organization for Standardization
LAC	Library and Archives Canada
MRA	Machine Readable Archives
NAC	National Archives of Canada
NARA	National Archives and Records Administration
NAS	Network Attached Storage
OAIS	Open Archival Information System
OCLC	Online Computer Library Center
OS	Operating System
PAC	Public Archives of Canada
PDI	Preservation Description Information
POWRR	Preserving (Digital) Objects With Restricted Resources
PREMIS	PREservation Metadata: Implementation Strategies
RAM	Random Access Memory
RLG	Research Libraries Group
SAA	Society of American Archivists

SIP	Submission Information Package
TRAC	Trustworthy Repositories Audit and Certification
UBC	University of British Columbia

Introduction

One of the major struggles for the archival community, especially over the last 30 years, has been the acquisition and preservation of born digital records. Born digital records are any records that were first created through a digital means. This could be something like a digital audio recording, pictures taken with a digital camera, or something more common such as a word processor document or e-mail.¹ What has happened since the 1980s, with the introduction of mainstream personal computing, and the 1990s, with the widespread introduction of the Internet, has been a greater use of digital technologies. This means that people, bureaucracies, governments, and corporations have come to rely more and more on digital means to create and manage their information. The problem for archives then has been a fundamental shift in how wider society creates and consumes information. We have moved collectively from a society with more traditional means of communication, such as paper-based correspondence, or oral communication by telephone or in person, to one increasingly reliant on the digital. With technologies like cell phones, computers and many internet based means like social media and e-mail becoming staples of how our society manages and spreads information archivists face the challenge of having to deal with an ever-increasing amount of born digital content. This has made digital preservation a vital area for archivists, one that is quickly becoming central to the profession. The Society of American Archivists (SAA) has recognized the centrality of digital preservation in its guidelines for graduate programs. They require that programs offered in archival sciences must have in

¹ Shelby Sanett, "Archival Digital Preservation Programs: Staffing, Costs and Policy," *Preservation, Digital Technology & Culture* 42, no. 3 (2013): 139.

their core archival knowledge requirements that students must be able to understand and work with digital records and access systems.² This shows that archivists recognize the central role these records take in our archival systems today.

Archivists need to continue their work with digital records because in many ways they have become fixated on the theory side of this issue, while more and more records are being created as born digital and the profession seems to lack the infrastructure and/or skills to deal with them effectively. We as a society have already felt that keeping our records preserved is important, as they provide a framework for accountability, research, and heritage. We risk losing a large portion of these records if archivists cannot successfully transition into digital archiving. Right now archivists seem to be stuck in a transitional phase into the digital. The process of digital preservation while being undertaken by archivists needs to become more central to the profession and something handled by all archivists.³ Some institutions have people who focus specifically as digital archivists, but that no longer seems enough in the face of how many records are now born digital. There will continue to be a place for those who work with legacy or traditional media as

² Ciaran B Trace, "Beyond the Magic to the Mechanism: Computers, Materiality, and What It Means for Records to Be "Born Digital"," *Archivaria* 72 (Fall 2011): 7.

³ There is a seemingly contradictory idea present in this thesis that does need to be unpacked. I claim that the work of digital preservation needs to be done, while also claiming that there are widespread digital standards. These ideas conflict in the sense that if digital preservation was not being done to the level it should be it does not stand that widespread standards would exist. But this is explained through there being many archives with many approaches to digital preservation, and by the fact that these standards exist in wider communities. For example OAIS started with the space industry's data preservation needs. Finally, while there are many archivists who are aware of and even endorse standards such as OAIS, this awareness and support does not always translate into implementation.

they do make up a large part of archival collections, but it will be necessary going forward for all archivists and archives to be able to deal with born digital content.

This is why the type of DIY, in-house solution I describe in this thesis is important. It is necessary at this point to move towards more action. Starting something like that can be difficult. Archival studies programs tend to pull heavily from people with Arts degrees. The knowledge and skills gained from Liberal Arts degrees make most archivists extremely capable to understand the theory and uses of archives, but it does not make for a wide pool of candidates who are well versed in computers and other digital technologies. It is rare for someone to come into the field of archival studies with a solid background in computer science or something similar. Thus many archivists have to develop their digital competency on the job, through supplementary courses and seminars or through their own work. This is why a DIY, in-house solution to digital preservation can be extremely useful. A DIY solution allows for archivists to work at their own pace and build up their skills, competency, and understanding as they work with digital technologies without becoming overwhelmed.

Similar to the DIY approach is what is outlined in the 2011 *Archivaria* piece by two professors from the School of Information and Library Science at the University of North Carolina, Chapel Hill Christopher Lee and Helen Tibbo. Their piece, titled "Where's the Archivist in Digital Curation? Exploring the Possibilities through a Matrix of Knowledge and Skills," cites the definition of digital curation by Elizabeth Yakel: "The active involvement of information professionals in the

management, including the preservation, of digital data for future use.⁴ Lee and Tibbo argue that more than archivists and archival theory are involved in digital preservation. For digital preservation to work archivists need to get outside their own field, and get others involved in their processes. Subjects like computer science, digital forensics, and records management are part of these outside areas that archivists would have to work with.⁵ This is similar to my proposed DIY solution, but more formal. The DIY idea I will describe later in this thesis would have archivists more involved with active digital preservation, which requires that they become involved in these other areas that Lee and Tibbo describe. This idea of digital curation is critical for digital preservation because archivists need to understand how digital records are created, used, managed, and the processes they will use in order to preserve them.

DIY as a solution for digital preservation plays into ideas that are advocated for by Ciaran Trace in her already mentioned article. Trace writes that “Understanding and working with born-digital material requires us to open the computer’s black box and become as familiar with computer hardware, application software, system software, and firmware as we have become with earlier writing, recording, and storage techniques.”⁶ Trace recognizes that the actual operations of hardware and software influence record creation. We have to understand these processes, and how they result in born digital records, so as to be better at

⁴ Christopher A. Lee, and Helen Tibbo, “Where’s the Archivist in Digital Curation? Exploring the Possibilities through a Matrix of Knowledge and Skills,” *Archivaria* 72 (Fall 2011): 124.

⁵ Lee and Tibbo, “Where’s the Archivist,” 124-26.

⁶ Trace, “Beyond the Magic,” 5

preserving them. She advocates for archivists to become a kind of computer hobbyist.⁷ Archivists should be active and engaged in a wide variety of new and legacy digital tools and techniques. This is similar to what I describe here as the DIY approach to digital preservation. Trace wants archivists to be active in continuously exploring digital technologies. This way they will not fall behind the technological curve, and will have a wide body of experience and literature to draw on in order to conduct digital preservation. Archival theory is foundational to Trace's thinking and to my DIY approach, but so is practical experience with digital technologies. These should not be mutually exclusive.

This attitude however is not entirely mainstream within the archival community. Rather than doing the work some want to outline the theory side or define the role of the archivist when born digital content is being neglected. Grant Hurley, in his 2016 *Archivaria* piece, quoted Trace on the meaning of born digital in the interplay of theory versus action. "As Ciaran Trace notes on the meaning of 'born digital,' archivists are trying to figure out what the appropriate nature and level of engagement should be with computer systems prior to jumping in to the deep end of developing necessary knowledge and infrastructure to actually do digital preservation as a routine activity."⁸ There is a frustrating attitude among archivists that there needs to be an agreement on the bare minimum of what an archive should be doing for digital preservation before starting the process. While this is an important question, archivists are losing time to learn the skills and start the work

⁷ Trace, "Beyond the Magic," 6-8

⁸ Grant Hurley, "Community Archives, Community Clouds: Enabling Digital Preservation for Small Archives," *Archivaria* 81 (Spring 2016): 130.

of digital preservation while waiting for that solution. These things can be done simultaneously. Archivists can work on the theory and engage with the records. Born digital records are the majority of what is presently created, and the backlog of born digital content will only continue to grow. Archivists do need theoretical guidance, but the longer digital records are left without intervention, the harder it is for them to be preserved in the future. This is why both methods should exist in a dialogic relationship. Archivists need to continue to explore and establish the theoretical backbone that will guide this work. At the same time, there has to be attempts to preserve current records and ongoing attempts to build systems and infrastructures for digital preservation today.

The challenges, needs and resources of every archival institution are unique. Geography, borders, language, financial resources, and other factors create great diversity among archival institutions. With that in mind it becomes harder to see digital solutions that can apply to all institutions. Instead it seems more effective and worthwhile for archives to put together their own solutions that address their needs and meet their challenges, with the aid of recognized international standards. This allows for an institution to account for their resources and needs, and plan an effective digital preservation strategy accordingly. The best way to set about creating such an individualized plan is to have the skill set and resources available for the archive and its staff to create this plan from the ground up. This is why I see DIY, in-house solutions as a very useful path forward for digital preservation. A solution of this type would be very much in line with what Ciaran Trace envisions in "Beyond the Magic to the Mechanism," where she suggests archivists need to take a

more active role as computer hobbyists.⁹ By taking a DIY approach, archivists and archives would develop the necessary equipment and skills for their digital preservation needs over time. As Trace suggests, this would better allow them to understand the hardware and software that create the targets of digital preservation efforts. This DIY solution then would let archivists work at their own pace, slowly building their skills and knowledge for digital preservation.

My archival education, including my year of archival studies course work, my internship and my thesis research, allowed me to slowly build up the background knowledge necessary to understand the history and theory of digital archiving that I describe in the first chapter of this thesis. I then move through a case study and analysis of the DIY in-house solution I developed at the University of Manitoba Faculty of Medicine Archives, under the guidance of then-Medical Archivist Jordan Bass, in the second chapter. In the third chapter I take three recognized international standards, which are covered in detail in chapter one, and use them to audit the workflow from the case study in Chapter Two.

The first chapter of this thesis is foundational to this topic, and is divided into four parts. First I look back roughly fifty years to write a brief history of digital preservation in Canada from the 1960s to 2011. I did this by doing my own primary source research on the published materials of the early digital archivists that did this work, but also with the help of some secondary work by the historians Betsy Baldwin and Greg Bak. These two have been extremely influential in this section as both have written their own histories of digital archiving in Canada and have given

⁹ Trace, "Beyond the Magic," 6-8

me a lot of information from their interpretation of primary sources and an idea as to what primary sources to seek out. The point of this part of the chapter is to establish the historiography and to show where some of my inspiration for the workflow has come from. There was an attitude of learning by doing during the mainframe-computing era when archivists were confronted with a new medium that they had to learn to process¹⁰. I think that while the archival community, and computer technology, has moved past that era, there is still a place for that mindset. The second part of this chapter details current digital standards that are extremely influential in the archival community. These are the Consultative Committee for Space Data Systems (CCSDS) *Open Archival Information System* (OAIS) reference model (International Organization for Standards (ISO) 14721), the PREservation Metadata: Implementation Strategies (PREMIS), and the auditing framework of Trustworthy Repository Audit & Certification of trustworthy digital repositories (TRAC) (ISO 16363). I describe the development of these standards and analyze their requirements to make a case for why they matter and are included in this thesis. In the third section of chapter one I lay out some current solutions for digital preservation, to establish some context for the environment in which this workflow was being made. The workflow I developed at the Faculty of Medicine Archives was conceptualized as an alternative to one of the solutions, Archivematica, so understanding it is key to following the workflow in chapter two. I conclude chapter one with a statement on where things currently stand for digital

¹⁰ Greg Bak, "Media and the Messengers: Writings on Digital Archiving in Canada from the 1960s to the 1980s," *Archivaria* 82 (Fall 2016): 76-7. Betsy Baldwin, "Confronting Computers: Debates about Computers at the Public Archives of Canada during the 1960s." *Archivaria* 62 (Fall 2006).

preservation in Canada. This is done by analyzing several journal articles that detail digital preservation projects from the last decade that give a sense as to what the current climate is like.

Chapter two builds on this foundation as I tell the story of my internship at the Faculty of Medicine Archives at the University of Manitoba in the summer of 2016. This chapter starts by explaining the situation at the archives that led Medical Archivist Jordan Bass and myself to look beyond Archivematica to create our own digital workflow. The chapter then moves from a narrative about this change to exploring the workflow itself. This is done in two parts. First, I provide a step by step explanation of how our workflow took a document from ingest to preservation and access. Secondly, I describe the tools that we used throughout the workflow. While there is some discussion of hardware, this section deals predominantly with the software that we used in the workflow, and offers a general idea as to how and why it was used. The final part of this chapter moves into a self-reflection on the workflow itself. This was a chance for me to critically engage with the workflow after I had created and used it as an archivist.

The final chapter draws upon content from the first and second chapters. In the first chapter I established three international standards that are influential in the archival community and have important metrics for long-term preservation of digital objects. The final chapter takes these standards and uses them to audit the workflow from chapter two. I do this by taking the international standards and comparing their requirements for compliance to the workflow. Working with OAIS, then PREMIS, and finally TRAC, I describe what each requires to be compliant with

it, and evaluate if the workflow meets that requirement. If it does not meet that requirement I have tried to establish a possible path towards resolving that. This chapter concludes with a general review of the audit and its results.

The purpose of this thesis is to establish the background knowledge to understand the case study in chapter two and present that case study as a possible path forward for digital preservation. Digital records are a pressing issue for the archival community. It is necessary that archivists continue to confront and resolve this challenge in different ways. Moving forward with a DIY solution as described in chapter two is a way to reincorporate the 'learn by doing' attitude of the past. Rather than waiting for others to solve digital preservation challenges, Archivists can follow Trace, and Lee and Tibbo, and use the work of experts from other fields to help solve this issue. A DIY solution will help archivists to build up their skills and assemble the often free and open source resources necessary to stabilize and work with digital records. I do not necessarily see this as a long-term solution, but it can help now while more formal theories and solutions continue to be developed. If we do nothing while waiting for these more formal solutions to arrive, we will drown in the digital deluge. DIY solutions that are tailored to the resources and mandates of each archives, and the abilities of each archivist, will help us to build capacity within our institutions, our profession and our selves, while also making real progress towards the digital archives that Canadian society needs for our digital age.

Chapter 1

Computers have, from the very origins of computing, been a key part of government work. In their early forms as large mainframes, computers were the perfect tool to automate processes like census and payroll calculations. This made them invaluable tools for government bureaucracies. This also meant that government records started to be produced by computers, in formats that were only able to be accessed by using a computer. This created a problem for archivists, since they lacked access to computing technology. Was it enough to simply print the results and preserve them? Or did the inputs also have to be saved? There was also the question of accompanying documentation, and how to arrange this data in traditional archival structures like fonds and how it would be possible to respect the original order with computer based records. In Canada at the federal level this issue caused the expansion of the mandate of the federal archives in September of 1973, when the Machine Readable Archives Division was established. The mandate of the division was to appraise, ingest, process and preserve records created by computers both in the federal bureaucracy but also in the private sector if they held national significance.¹¹

To trace the history of digital archiving in Canada, I begin with how the Machine Readable Archives (MRA) Division was established, and explore the progress of events and ideas surrounding digital records up to the most recent “digital archives” special issue of the Association of Canadian Archivist’s journal *Archivaria* in 2011. In the second part of this chapter, I will look outside of Canada to

¹¹ Baldwin, “Confronting Computers,” 159–78

consider several key standards from the 1990s onward that have become central to the archival community's ideas about digital preservation. These are the Open Archival Information System (OAIS), PREservation Metadata Implementation Strategies (PREMIS), and the Trustworthy Repositories Audit and Certification: Criteria and Checklist (TRAC). The final part of this chapter evaluates some current digital solutions that are available to archivists in Canada and ends with a general statement of the current state of digital preservation in Canada.

My own formulation of this history of digital preservation, in this first section, has relied heavily on two scholars and their work on similar projects that traced the history of digital preservation in Canadian. Betsy Baldwin, in her doctoral thesis and a later article in *Archivaria*, traced the development of the MRA Division and digital preservation at the federal level from the 1960s to the 1980s. Greg Bak also wrote two articles that traced the history of digital preservation in Canada. "Media and the Messengers" traced the development of digital preservation from the 1960s to the 1980s telling the story of the MRA, but also the larger conversation of digital preservation within Canada at that time. "How Soon is Now?" continued the story by looking at how digital preservation advanced without the MRA, to the attempts at total digital repositories in the 1990s and ending with the second *Archivaria* special issue on digital preservation in 2011.

Before the establishment of the MRA Division at the Public Archives of Canada (PAC), 1973, there was some experimentation with computers and digital preservation. The most notable example is that of an early project in the 1960s to use computers to help generate finding aids. During this period the largest division

at the PAC was the Manuscript Division. From 1959 to 1966 they saw research requests increase from 1314 to 4363, and the archival records they were responsible for increased from 315 feet of records to 4223 feet.¹² To respond to this increased work load a group within the Manuscript Division began to experiment with the possibility of using computers to help generate finding aids for the Prime Ministers' papers collection. Jay Atherton, a key member of this project, went on to write a paper on the possibility that automation presented for archives. Published in 1970 but based on a paper first presented in 1968, "Automation and the Dignity of the Archivist" makes the argument that computers were a natural fit for automating the creation of finding aids. Using computers would limit human errors and their function of tabulating and organizing data quickly would speed the creation of finding aids.¹³ There were, however, challenges with this project. Computers were not sophisticated enough to deal with the multiple ways of entering data for both search and reference terms. For example, the computers could not recognize that John A. Macdonald and Macdonald, John were the same person.¹⁴ This project ultimately went over deadline and budget, by roughly \$14,000 and a year.

The year 1968 marked another key shift for the institution when Dr. Wilfred Smith replaced Dr. William Lamb as Dominion Archivist. This change in leadership saw two major shifts that started the institution towards the creation of the MRA Division. First, in 1969, Smith sent Michael Carroll to the United States National

¹² Baldwin, "Confronting Computers," 162.

¹³ Atherton, Jay. "Automation and the Dignity of the Archivist." *The Canadian Archivist* 2, no. 1 (1970): 56-58.

¹⁴ Betsey Baldwin, "Stepping off the Paper Trail? Rethinking the Mainframe Era at the Public Archives of Canada," (Ph. D History-Specialization in Canadian Studies), University of Ottawa, 2006), 100-102.

Archives and Records Service to study and report back on the American's machine readable records program that was underway at the time. The second major event took place in 1971 when the International Council of Archives launched the study "Automation and Archives," under French archivist Robert-Henri Bautier, who felt that archives around the world lacked computerization. Bautier's fear was that if archives failed to address computerized tools and records, traditional archives might find themselves displaced as repositories for digital records by electronic data processing centers and university data centers. Carroll expressed this same fear to students in the PAC education programs in 1971 and 1972.¹⁵

In 1972, partly in response to these concerns, and a request from the Treasury Board for the archives' input on a report on electronic data processing management, the Public Archives founded a committee, including both Carroll and Atherton, to give an archival perspective on the management of machine readable records. The result of this committee was that PAC submitted their "General Policy Statement re the Role of the Public Archives of Canada in the Field of Automatic Data Processing," to the Treasury Board on June 1, 1972. In this report, they concluded that the Public Archives Act of 1912 and Public Records Order of 1966 gave the Public Archives considerable authority for the care and condition of public records still in use by departments of the federal government. Their ultimate desire, stated in this report, was authority over automatic and electronic data processing record creation in federal departments and agencies, including the authority to inspect systems and processes, and to regulate record retention, storage and

¹⁵ Ibid, 119-20.

disposal conditions. They also recommended that they be allowed to ingest, store and preserve all dormant automatic data processing records and manage any that could be deemed historically valuable as part of their collection.¹⁶ Further strengthening this position was the 1972 acquisition, overseen by Michael Carroll and Sue Gavrel, of the punch cards and code books created by the Royal Commission on Bilingualism and Biculturalism.¹⁷

The Treasury Board took this proposal quite seriously and asked for a fuller articulation, financial estimates and a possible implementation plan for this initiative. This marked the true beginning of the MRA Division in the Public Archives. The official approval came a year later in the summer of 1973 when the Privy Council approved the machine readable records program by giving the Public Archives two new areas of responsibility: scheduling of electronic data processing records in all federal departments and agencies, and the establishment of a machine readable archives division to preserve and service inactive records of national historical value.¹⁸

Two years after this Michael Carroll published an article discussing the early experience that the PAC had while they established the MRA Division. The ultimate objective of the program that created this division was to provide archival services for machine readable records that could be considered to have archival value and were produced by the federal government and the private sector, if they held some kind of national significance. Carroll explained that they interpreted this objective as

¹⁶ Betsey Baldwin, "Stepping off the Paper Trail?," 120-31.

¹⁷ Ibid, 126-7.

¹⁸ Betsey Baldwin, "Stepping off the Paper Trail?," 130-131 and 139-141.

meaning that they had to: acquire machine readable records of long term value, make sure these records were properly preserved, and ensure that they could provide proper reference services for users.¹⁹ This represented a change in medium from traditional archives. Archival theory had previously been based around physical objects and textual records. Archivists were trained and worked in an environment that reflected this, by working to acquire, appraise, and preserve textual records on paper or some similar medium like vellum. There were archival objects outside textual records that archivists had to preserve, for example, PAC had at this time units for the preservation of images, maps, audio/visual and other record types. However computer records were a new medium that presented unique preservation challenges.

Carroll explained that to reflect this difference in medium that computer based records presented, they changed their perspective to meet this new medium, rather than sticking to the traditional criteria of archival value. No longer could archivists wait for a records life span to end, since magnetic tape, the principal medium for machine readable records at this time, is extremely vulnerable to degradation or deletion. Getting involved earlier, while records were still in use, also made it more likely that archivists would be able to acquire the supporting documentation that provided the context necessary to interpret and use these machine readable records.²⁰ Once acquired, the MRA chose to use magnetic tape for storage of the records as the cheapest and easiest option. Although they did

¹⁹ Michael Carroll, "The Public Archives of Canada's Experience in Establishing a Machine Readable Archives," *The Canadian Archivist* 2 no. 5 (1974), 53-4.

²⁰ Carroll, "The Public Archives of Canada's Experience in Establishing a Machine Readable Archives," 55-7.

recognize that magnetic tape was not reliable over the long term, they addressed this by keeping two copies on tape and one copy on computer output microfilm. For their reference services they chose to copy tapes and send digital copies out rather than provide access to computers and data processing staff in the reading room.²¹ Through Carroll's work we can see the early digital preservation processes and how they encourage ideas that are still used today, like pre-custodial intervention and preservation through multiple copies (i.e. redundancy).

During its first five years the MRA Division was fairly successful in its attempts at preserving computer records. In total they appraised over 2500 data files, acquiring more than 500, and served around 250 researchers. However, there were major issues with acquisition of two collections from federal departments that used computers. Statistics Canada had one of the oldest and largest sets of computer based records in the federal government. The Anti-Inflation Board also heavily relied computers to process information concerning individuals and businesses within Canada. Both departments refused to transfer, and the Anti-Inflation Board even expressed their intention to never transfer their computer records citing federal legislation and concerns over the personal information these records held. The *Statistics Act* and *Income Tax Act* both outlined how these departments were to collect and manage this information. The *Statistics Act* was interpreted by Statistics Canada as forbidding them from transferring their records to the Public Archives and The *Income Tax Act* allows only for the Finance Minister and members of the Anti-Inflation Board to view their records and does not specify that their records

²¹ Ibid

can be viewed by the Public Archives to be managed as public records. In the end these departments destroyed their records without review from the Public Archives.²²

As well as struggling to obtain certain documents the MRA had difficulties in attracting users for the records they did preserve. This lack of interest in their records did affect them when in 1985 when development started on a new National Archives Act. This act introduced a 9% cut to the federal archives budget for staff, and Dr. Smith was replaced by Dr. Jean-Pierre Wallot as Dominion Archivist. Major restructuring of the Public Archives in 1986 resulted in the closure of the MRA Division through its amalgamation with the Federal Archives Division into a new Government Archives Division.²³ The Public Archives then went from being a national and worldwide leader on machine readable records over the 1970s and early 1980s to fizzling out and ending by 1986. In the 1980s there was a fundamental shift away from mainframe and mini-computing, which had been the focus of the MRA, to desktop computing. Bak argues that archivists fell into the common pattern of hyping the machine itself as a revolution, instead of seeing desktop computing as a continuation of mainframe and mini-computing. As record creators and archivists adopted the cheaper computers of the desktop era digital records became even more of a concern for contemporary archivists. However, this shift in computing culture did not see archivists attempt to turn to the already

²² Betsey Baldwin, "Stepping off the Paper Trail?" 172, and 189-91

²³ Ibid, 251-2

established literature and knowledge of the MRA era to deal with digital records.²⁴ Digital preservation did continue, but without the consultation or guidance of experienced and knowledgeable Canadian MRA archivists. As Bak says, desktop computing appeared to come out of nowhere for those working with analogue records. This fundamentally ahistorical interpretation essentially erased the long and established history and work of the MRA experts.²⁵ The closure of the MRA Division at PAC in 1986 shifts the story of digital archiving to one that can be explored through the publications of the few digital archivists in Canada at the time, including those who had previously served in the MRA Division.

A year after the closure of the MRA Division Hugh Taylor, an influential archivist at PAC during the 1970s, argued that the new medium of digital records presented such a drastic shift that archivists needed to rethink their craft. This argument was done with an analogy to Transformer toys, saying that archives used to be reminiscent of jigsaw puzzles in that they would be set out in a pattern ordered by concepts like the *fonds* or original order.²⁶ According to traditional archival theory, archives, like a jigsaw puzzle, had only one solution possible, one place for each piece that was determined by the shape of the piece and what it fit next to. To Taylor, digital records are like Transformers in that they change their shape and form or revert back, but are still made up of the same original pieces.

Taylor used this analogy to draw attention to something that archivists still struggle

²⁴ Greg Bak, "Media and the Messengers: Writings on Digital Archiving in Canada from the 1960s to the 1980s," *Archivaria* 82 (Fall 2016): 57-9.

²⁵ Greg Bak, "How Soon Is Now? Writings on Digital Archiving in Canada from the 1980s to 2011," *The American Archivist* 79, No. 2 (2018): 287

²⁶ Hugh Taylor, "Transformation in the Archives: Technological Adjustment or Paradigm Shift?," *Archivaria* 25 (Winter 1987-8): 12-15.

with today: how do we as a profession arrange and present digital records when the possibilities are, in theory, limitless. Computer based records defy the conception of original ordering and *respect des fonds* as it exists for physical records. It was hard for archivists at the time to envision that they could recreate original order for digital records when their frame of reference was for physical records, which exist as unique originals that can be physically arranged in only one order, rather than digital copies on a hard drive, punch card or magnetic tape, that can have many digital representations and orderings. Taylor's analogy of the Transformer toy, then, attempted to show that archivists might move beyond original order and the fonds system in archiving digital records, presenting digital data in new ways, emphasizing new or different themes, and allowing collections to intermix and cross over. Taylor's proposal represented a fundamental shift that the profession has yet to work out.

Taylor and his work drew attention to the challenge of the digital objects that archivists were attempting to preserve. However, this is not the entirety of digital preservation. Propriety hardware and software creates a new environment for records wherein the record and the ability to access it both have to be preserved. In 1986 John Mallinson wrote an article discussing this concern for the preservation of machine readable records as the director of the Center for Magnetic Recording Research at the University of California. His central point was that preserving the records is unimportant if the machines necessary to access the records fall out of use and disappear. His worry was that in the 1980s devices were meant to be used

for no more than 20 years, with 10 years being a reasonable period of use.²⁷ As with Taylor's observations on original order, Mallinson identified a key challenge that has not yet been resolved.

In the 1990s, changing computing technology continued to upset the world of archivists trying preserve the computer based records of Canadian society and government. In 1995 John McDonald, a former member of the MRA Division, summarized the current landscape of digital technology as a "wild frontier." The analogy of the wild frontier was apt because it accurately described the individual freedom created through technology to easily create, copy, store, and delete records. This made the Internet and technology like the wild west, where rugged individualism was the key to thriving.²⁸ But civilization and order were creeping in, represented by the push towards integrated software and specific distinct tools that supported the automation of work processes. He noted that the ease of creating and sending an email circumvented processes of review and approval. McDonald asked some of the basic questions that have shaped digital preservation to this day, such as if an email is a record, and if so what to do with it. In navigating new technologies and new corporate cultures it was necessary figure out how to acquire and catalogue the right digital objects the right ways, and in context.²⁹ McDonald's analogy suggested that, as record keeping regulations encroached on the "wild west" of the digital office of the 1990s, archivists and record managers had the

²⁷ John Mallinson, "Preserving Machine-Readable Archival Records for the Millenia," in *Archivaria* 22 (Summer 1986): 147-50.

²⁸ John McDonald, "Managing Records in the Modern Office: Taming the Wild Frontier." *Archivaria* 39 (1995): 70-79.

²⁹ *Ibid*, 71-3.

opportunity to influence how records were created, stored, and transferred to archives or deleted.

Some take issue with his conceptualization of the state of digital archiving at the time. In “How Soon is Now,” Bak critiques McDonald’s analogy, drawing on archivist Margaret Hedstrom. Bak argues that McDonald’s “wild west” analogy is problematic as it is ahistorical and colonial in nature, painting the pre-American West uncivilized when it was in fact inhabited with Native American civilizations. For Bak and Hedstrom, the analogy is further reductive for two reasons. First, McDonald’s characterization of the earlier period as some kind of anarchy minimalizes, if not erases, the contributions of MRA archivists from the history of digital archiving. Second, McDonald misrepresents the reality of what was taking place in regards to digital preservation. Hedstrom argues that digital record management in the 1990s became more about cooperation and collaboration between office workers, IT professionals and record keepers.³⁰

In the 1990s there were three large electronic record projects in North America that tried each in their own way to deal with the growing mass of born digital records. The first two projects, which became extremely influential in digital archiving, were the 1993 to 1996 Pittsburgh Project overseen by David Bearman and Richard Cox, Bearman being an influential voice on digital preservation in the worlds of museum and archival theory and Cox a professor at the Pittsburgh School of Information Sciences, and the 1994 to 1997 University of British Columbia (UBC) Project run by Luciana Duranti, a professor at the UBC archival studies program,

³⁰ Bak, “How Soon Is Now?,” 292-3.

with Terry Eastwood, the founder of the UBC program. The Pittsburgh Project rejected past approaches, such as those of the MRA Division, as being focused only on outputs. Instead, Bearman and Cox advocated for working from conceptual models that would locate essential information within the live system itself. Bearman felt that one could not remove data from the recordkeeping system as this would undermine the archival properties of the data, which could not be properly interpreted and accessed without the system.³¹ Duranti by contrast focused her project around the concept of diplomatics, through which records are shown to be authentic through their content and contextual information.³² The third major project of the 1990s, was at the National Archives of Canada (NAC) and was called Information Management and Office Systems Advancement (IMOSA). This was an attempt to set up a digital environment that office workers would work in that forced them to be consistent with recordkeeping principles.³³ This was an attempt at the early intervention that earlier digital archivists like Atherton and Carroll had argued for in the 1970s. It was a heavy-handed attempt to control the way records were created rather than to document the digital technologies in use.³⁴

These three projects looked towards radical solutions that would have transformed the digital office to make it compliant with record keeping objectives. Contemporary digital cultures, however, remain focused on operational needs rather than record keeping requirements, with some limited exceptions. Rather than seeking to reengineer records creation environments, digital archiving and digital

³¹ Bak, "How Soon Is Now?," 283–319.

³² Ibid, 295-304

³³ Ibid, 297

³⁴ Ibid, 297-8

preservation in the late 1990s and early 2000s came to focus on broad, open, and collaborative projects like the OAIS.

Digital Archival Standards:

There are several key standards that pertain to digital archiving. However, there are three that form the foundation of most digital solutions and best practices for digital archivists today. These are the Consultative Committee for Space Data Systems (CCSDS) *Open Archival Information System (OAIS)* reference model (International Organization for Standards (ISO) 14721), the PREservation Metadata: Implementation Strategies (PREMIS), and the auditing framework of Trustworthy Repository Audit & Certification of trustworthy digital repositories (TRAC) (ISO 16363). These provide the language and methodological base for contemporary collection and ingest of digital records, as well as a system to understand the necessary metadata to preserve a record, and a system through which to audit an institution as trustworthy of holding and managing digital records.

OAIS was developed in the 1990s under the leadership of the CCSDS. Formally their area of expertise is space agencies; they convene periodic meetings to address data system problems common to all members. Through this effort they hoped to devise possible solutions to these problems. In the 1990s many record keeping professionals called for a more uniform way to do digital archiving, but it took members of the space research community to actually drive this home through their development of OAIS.³⁵ This standard went beyond their industry and has become one of the pillars of archival theory supporting digital preservation. There

³⁵ Christopher A. Lee, "Open Archival Information System (OAIS) Reference Model," *Encyclopaedia of Library and Information Sciences*, (2010): 4020-3

are two key aspects to note about OAIS before one can understand it. First, participation in the CCSDS is voluntary. OAIS makes recommendations only because as a voluntary organization these cannot be binding. Second, it is important to clarify that while the O in OAIS stands for open, this is not to suggest that the archive must be open or use open technologies, but that OAIS was developed in open forums.³⁶

An OAIS archive is defined as an institution that intends to preserve information for access and use by a Designated Community.³⁷ Information is defined as any kind of knowledge that can be exchanged, and is always expressed by some type of data in an exchange. This exchange also requires the exchange of Representation Information, which is something that is needed to understand the received information. The example given in the OAIS Magenta Book is that one may need a dictionary to understand a language and therefore to be able to read a text.³⁸ The dictionary is the Representation Information, and the text would be the information being exchanged in this example. In digital terms this could be expressed as bits being the information and the necessary hardware and software to render said bits as useable would be the Representation Information. But the point is also made that an OAIS institution must understand the Knowledge Base of its

³⁶ Christopher A. Lee, "Open Archival Information System (OAIS) Reference Model," 4021

³⁷ In OAIS several key terms are capitalized to denote that terms specific definition as outlined by the CCSDS is being used. I have chosen to follow this same convention when using terms as they are defined in OAIS.

³⁸ CCSDS Secretariat, *Reference Model for an Open Archival Information System (OAIS)*. Washington, D.C.: CCSDS, 2012, 1-14 to 1-15.

Designated Community.³⁹ This allows for it to provide the proper minimum Representation Information so that said community can understand the information the archives is maintaining. So then to properly preserve information in an OAIS approved way it is crucial to clearly identify and understand the Data Objects one is preserving and the associated Representation Information. Without this Representation Information (i.e. a dictionary, in the example) it is possible that the Designated Community would not be able to properly utilize the preserved information.

OAIS gives its recommendations in two forms: Recommended Standards and Recommended Practices. Standards are prescriptive and serve as the method to specify how infrastructure should operate and interoperate. Practices are more descriptive and are intended to provide general guidance about how to approach particular problems associated with preservation.⁴⁰ These practices are issued by and represent the consensus of the CCSDS members, so these practices remain voluntary. There is no commitment by any agency or organization to implement these Recommended Practices and instead they are intended to serve as guides. Practices are reviewed every five years to decide if changes in technology, requirements and/or directions are warranted. The point of this reference model then is to provide a framework for understanding and increasing awareness of the concepts necessary for long-term preservation and access of digital information. OAIS supports this by: providing the concepts needed by archival and non-archival organizations to be effective participants in the preservation process; providing the

³⁹ Ibid, 2-3

⁴⁰ CCSDS Secretariat, *Reference Model*, ii

framework for describing and comparing different long term preservation strategies and techniques, and defining the important elements and processes for long term digital preservation and access, among other important areas related to long term preservation and access.⁴¹

OAIS is able to provide broad support for digital preservation and access to information through the wide range of archival functions it addresses related to preservation. These functions include ingest, storage, data management, access, information dissemination, the migration of digital records to new media and formats, the necessary role of software for preservation and access, and the exchange of digital information using archives. By providing loose roles around these key archival functions the standard helps institutions put in place long-term preservation and access.⁴² This has proven to be a key strength of OAIS. Its widespread influence rests upon its open and accommodating nature. Its authors state that they make no assumptions or endorsements of any specific computing platform, system environment, design paradigm, development methodology, database management system, database design paradigm, data definition language, command language, system interface, user interface, technology, or media.⁴³

Being OAIS compliant, is limited to supporting the model of information in section 2.2 and fulfilling the responsibilities listed in section 3.1 listed in the Magenta Book Reference Model by the CCSDS. Section 2.2 has three subsections that address how an institution will handle and define key concepts around their data to

⁴¹ CCSDS Secretariat, *Reference Model*, 1-1 to 1-2.

⁴² Ibid, 1-3 to 1-4

⁴³ Ibid, p. 1-1 to 1-2

properly preserve it. The first section deals with the definition of information, defined as any type of knowledge that can be exchanged and always expressed by some type of data in an exchange. This section also deals with Representation Information. The point of this section is that institutions need to recognize that for their users to get information from an archival institution three other things are required, two of which the institution must provide and the third they must account for. The institution needs to provide the Data Object (the records) and the Representation Information (the necessary contextual knowledge so the Designated Community can understand the Data Object), and must account for the Knowledge Base of their Designated Community so they provide the right Representation.⁴⁴ The combination of these three elements yields the Information Object.

The second section of OAIS accounts for how this data set is stored and presented, in something called an Information Package. An Information Package is a container that holds two types of information: Content Information, and Preservation Description Information (PDI). Together these two elements are represented by the Packaging Information, which is the information that is used to bind and identify the content of the package as well as make it accessible through the PDI. The Content Information is the information that was the original target of preservation and the Representation Information is the additional information necessary to understand it. The PDI can only be created after the Content Information has been clearly defined and assessed as the PDI is the information that is necessary to identify the Content Information, and to understand the environment

⁴⁴ CCSDS Secretariat, *Reference Model*, 2-3 to 2-5

it was created in.⁴⁵ OAIS divides PDI into the following five categories: Provenance, Context, Reference, Fixity, and Access Rights. Provenance displays who has had custody of the Content Information since its origin and throughout its history including the processing history at archival institutions. Context describes why the Content Information was produced and could include reference to other related Content Information. Reference provides one or more unique identifiers for the Content Information, such as an ISBN for a book. Fixity provides some type of check against undocumented alteration, such as a checksum. Access rights provide terms for access, preservation, distribution, and usage of the Content Information.⁴⁶

Section 3.1 lays out a list of six mandatory responsibilities that an institution must fulfill to be considered OAIS compliant. The first responsibility is for institutions to negotiate for and accept appropriate information from Producers, although what constitutes appropriate information is left up to the discretion of the institution and would vary depending on mandate and what is seen as necessary Representation Information. The second is to obtain sufficient control of the information to ensure long-term preservation. This falls under the legal ownership of a document. An institution has to secure proper legal authority to undertake key preservation actions, as well as to provide sufficient access, if they are to properly preserve and provide access to a record indefinitely. Thirdly, the institution must determine who their Designated Community will be and as part of this determine the Knowledge Base of the community so the proper Representation Information can be planned for and obtained. Fourth, they must ensure preserved information is

⁴⁵ Ibid, 2-5-2-7.

⁴⁶ CCSDS Secretariat, *Reference Model*, 2-5-2-7.

independently understandable to the Designated Community. Fifth, they must follow documented policies and procedure to ensure that their records are preserved against all reasonable contingencies. Finally, they must make their preserved information available to the Designated Community, ensure that information can be disseminated to Consumers and have evidence to prove their records as authentic. These are the six conditions that must be met for an institution to be considered OAIS compliant.⁴⁷

Part of the strength of this system is that it simply provides a set of rules that help to ensure that accurate records are preserved as long as necessary. They are sufficiently open ended that institutions can pursue their own individual preservation strategies, but still have some guidance on how to do digital preservation properly through these rules to ensure authentic data is collected and is able to be understood by a Designated Community.

One area that OAIS does touch upon, but avoids finer details is metadata. Representation Information includes metadata for both preservation and description. Representation Information exists to help describe the data object and to help preserve and interpret it. OAIS describes types of metadata without exploring the issues related to necessary metadata. A standard that was designed to build upon OAIS by addressing this need for proper metadata is PREMIS.

The *PREMIS Data Dictionary*, distributed by the Library of Congress, is promoted as a comprehensive and practical resource for implementing preservation metadata in digital preservation systems. Preservation metadata is characterized by PREMIS

⁴⁷ CCSDS Secretariat, *Reference Model*, 3-1.

in the following four ways. Firstly, metadata supports the viability, renderability, understandability, authenticity, and identity of digital objects. Second, it represents the information that repositories need to know for long-term digital preservation. Third, PREMIS emphasizes that metadata must be implementable, meaning that it is rigorously defined, supported by guidelines, and oriented towards automated workflows. Finally, in PREMIS metadata is defined as embodying technical neutrality. Therefore the system to collect, present and store preservation metadata should not have any assumptions made about preservation technologies, strategies, storage, and other factors. Like OAIS, PREMIS does not dictate technological details, but rather develops a larger picture of what preservation metadata is necessary for long-term preservation.⁴⁸ PREMIS originally came about from a collaboration between the Online Computer Library Center (OCLC) and the Research Libraries Group (RLG) in June of 2003. Their goal for this project was to bring together international experts in the use of metadata so as to help support digital preservation activities with metadata. The PREMIS Data Dictionary, the key publication for PREMIS, builds upon the OAIS reference model, using its definitions and structures of information objects and packages for the structure of the associated metadata. The data dictionary was meant to serve as a translation of the framework provided by OAIS into a set of implementable concepts. The key way that this would be done would be through the implementation of a preservation metadata scheme.⁴⁹

⁴⁸ PREMIS Editorial Committee. *PREMIS Data Dictionary for Preservation Metadata*. 3rd ed. Library of Congress, 2015, 1.

⁴⁹ *Ibid*, 1-2.

The model laid out by PREMIS defines four entities as important in regards to digital preservation activities: objects, events, agents, and rights. Objects are the units of information that will be subject to digital preservation. Objects include environments, which are defined as the elements of the technical stack (hardware, software, and other objects) necessary to render the representations, files and bitstreams. The category of objects is further divided into four subcategories. First is an intellectual entity, which is a distinct creation that is considered relevant to the Designated Community of the institution. Second is a representation. This is the set of files including the structural metadata needed to completely render and access an intellectual entity. Third, is a file. This is defined as a named and ordered sequence of bytes that is known to an operating system and that has a file format, access permissions, and file system characteristics. Finally, there is the bitstream, which is contiguous or non-contiguous data in the file that has meaningful properties for preservation purposes.⁵⁰ The second entity, events, is any action that involves objects or actors known to the system. This includes any action or person that affects the objects. These must be recorded in preservation metadata to show that there has been interaction or possible change. This is meant to allow for the proper establishment of provenance.⁵¹ Agents are the people, organizations and software that will be interacting with the object. Again these have to be recorded to establish proper provenance of the record.⁵² Rights as an entity pertain to the legal rights

⁵⁰ PREMIS Editorial Committee. *PREMIS Data Dictionary*, 6-7.

⁵¹ *Ibid*, 7

⁵² *Ibid*, 7

surrounding the record, determining who can make decisions about access and preservation for this record.⁵³

Focusing on these four elements and identifying relationships between objects allows the creation and collection of preservation metadata. PREMIS defines two types of relationships that are important to this: structural relationships and derivation relationships. Structural relationships show the relations between parts of objects and between files that constitute a representation of an intellectual entity. This is an essential piece of preservation metadata because without having a record of the structural relationships it would not be possible to understand and preserve the representation of a digital entity.⁵⁴ Without this understanding the original order of these files would not be preserved, and there may be lost information on relationships that may be necessary for long-term preservation. Derivation relationships result from the replication or transformation of an object. In this instance the intellectual content of the object remains the same, but its representation and/or its format will change. For example if I have file A in format X, but migrated that file to create file B in format Y, file A and B would then have derivation relationship. This allows format migrations to be represented through the two files and the accompanying preservation metadata that establishes the derivation relationship. While we may think of these as being the same object, under PREMIS it is not possible to change a file. Moving the file to a new format will fundamentally change the bitstream and representation, and thus create an entirely new digital object. Therefore these files A and B in the example should be thought of

⁵³ Ibid, 7

⁵⁴ PREMIS Editorial Committee. *PREMIS Data Dictionary*, 19

as two distinct objects with a derivation relationship created by the transformation event.⁵⁵ There is also a third relationship type that is more situational to environments. It is called a dependency relationship and it exists when an object requires an environment or a specific piece of hardware or software to support its function, delivery, and/or the coherence of its content.⁵⁶

PREMIS is intended to be as flexible as possible in its implementation. Like OAIIS there were no assumptions made regarding the nature of digital preservation systems in which it would be implemented, the preservation strategy being followed, or even the metadata management system responsible for creating the preservation metadata. Institutions must recognize the importance of those four elements and three types of relationships and have them represented in their preservation metadata. PREMIS recognizes the importance of describing all actions taken in preserving a record, including those of record keepers.

A third key standard for contemporary digital archival theory is Trustworthy Repositories Audit & Certification (TRAC) ISO 16363. This was designed to be a way for an archive to show that it is trustworthy and able to properly preserve and manage digital records. In 1996 the Task Force on Archiving of Digital Information declared that a critical component of digital archiving infrastructure is the existence of trusted organizations capable of storing, migrating, and providing access to digital records. They saw the major issue being that there was no formal way for an institution to declare itself as a trustworthy organization. There was a need to create a climate of trust about the prospects of preserving digital information, and at the

⁵⁵ Ibid, 19-21

⁵⁶ Ibid, 20

time there was no formally organized digital preservation community with common, consensus-driven practices or standards. Each repository conducted digital preservation in its own manner and to the level that addressed the funding and needs of their community.⁵⁷ The first trend towards any kind of movement of archives showing they were trustworthy was that many would declare that they were OAIS compliant to show that they were following some kind of standard on digital preservation that could be trusted.⁵⁸ In 2002 things started to change when the RLG and OCLC jointly published *Trusted Digital Repositories: Attributes and Responsibilities*. This report was meant to further develop a framework of attributes and responsibilities for trusted, reliable, and sustainable institutions that would be able to handle the wide range of digital records held by both large and small heritage and research institutions.⁵⁹

RLG and OCLC made this framework of best practices loose enough to be accommodating to a variety of institutions, but also made sure that it could still provide a basis for the expectations of a trusted repository. This 2002 report proved to be a useful tool, but it took combining it with other standards to make it effective, as it was just a verification process for trustworthy institutions, rather than a set of practices like OAIS.⁶⁰ The next year, in 2003, the RLG and the National Archives and Records Administration (NARA) created a joint task force on digital repository certification. The goal of this task force was to create criteria to identify digital

⁵⁷ Robin Dale and Bruce Ambacher, "Trustworthy Repositories Audit & Certification: Criteria and Checklist," Center for Research Libraries, 2007, 1-4

⁵⁸ However, it should be noted that OAIS was circling in drafts as early as 2000. Lee, "Open Archival Information System," 4022

⁵⁹ Dale, and Ambacher, "Trustworthy Repositories Audit," 1-4

⁶⁰ Ibid, 1-7

repositories capable of reliably storing, migrating and providing access to digital collections.⁶¹ This was challenging, as the criteria needed to work for a wide variety of heritage institutions from national to local. With massive differences in resources and skills it was difficult to make a system that would work for all levels of institutions.

In 2007, the Center for Research Libraries (CRL) and OCLC published the *Trustworthy Repositories Audit & Certification: Criteria and Checklist*. This was meant to represent current best practices for the organizational and technical infrastructure required to be considered a trustworthy institution capable of earning certification as such.⁶² At the base level a trustworthy digital repository must start with “...a mission to provide reliable, long-term access to managed digital resources to its designated community, now and into the future.”⁶³ The report provides criteria that trustworthy repositories should be able to meet, and provides explanations and examples to further explore and unpack these concepts.⁶⁴ To start with they unpack the idea of trust, suggesting that it goes beyond digital preservation systems and management of digital records. To determine trust requires looking at the entire system in which the digital information is managed, including the organization running the repository, its governance, organizational structure and staffing, financial condition and long term sustainability of the repository, and the contracts, licenses and liabilities under which it must operate.⁶⁵

⁶¹ Ibid, 2

⁶² Dale and Ambacher, “Trustworthy Repositories Audit,” Foreword

⁶³ Ibid, 3

⁶⁴ Ibid, 3-4

⁶⁵ Ibid, 3-5

This requires a deeper look than simply at how a repository stores records to deem it trustworthy. CRL and OCLC recognized that digital archiving does not take place within a vacuum, and that these other factors play a major role in determining if an institution can be considered trustworthy. It is not possible to just take the word of an institution that it is properly preserving records. Having some type of third party authentication that can inspire trust in a repository and its processes is a major reason why TRAC is important.

The types of issues that might make a repository untrustworthy include: the possibility of media, hardware or software failure, communication errors, failure of network services, media, software and hardware obsolescence, operator error, and others. To be considered trustworthy an institution must have a means of constant monitoring, planning, and maintenance, and take into account some kind of actions and strategies they can implement to help to prevent these issues.⁶⁶ TRAC certification is difficult and expensive to achieve and maintain.

CRL and OCLC ensure that users and auditors take into account the context of the institution itself when deciding if it is trustworthy or not. This includes considering the institution's mission, priorities, and stated commitments.⁶⁷ To provide direction on this they give four principles for applying criteria: documentation, transparency, adequacy, and measurability. Documentation is evidence that shows the objectives, design, specifications, and implementation of long-term digital repositories. This is something that should be documented and this documentation should be reviewed and updated on a regular schedule.

⁶⁶ Dale and Ambacher, "Trustworthy Repositories Audit," 3-4

⁶⁷ Ibid, 3-5

Transparency is another key factor when examining a repository for trustworthiness, and is considered internally and externally. Only when a repository exposes its designs, specifications, practices, policies and procedures for risk analysis can it truly be trusted. Adequacy simply means that at the most basic level an audit should answer whether an institution can meet its stated commitments. Finally, measurability means that the institution has objective controls or criteria against which it can be evaluated.⁶⁸ An institution able to meet these criteria could be certified as TRAC compliant and would have the standard's backing for its claim to be a trustworthy digital repository.

These three digital standards -- OAIS, PREMIS, and TRAC -- have come to form an important part of archival approaches to digital preservation. TRAC provides a framework for an institution to be certified as trustworthy, which is extremely important when dealing with digital records. However, while TRAC provides a framework it is not the final voice on determining the trustworthiness of an institution. That instead falls to the users of that institution.⁶⁹ As Duranti noted in the 1990s digital records are easily manipulated and changed and so there needs to be some element of trust and transparency for users of digital repositories.⁷⁰ PREMIS aids archives in setting up a preservation metadata scheme that helps to ensure that their records will be preserved and authentic for as long as necessary. OAIS provides the conceptual framework to ingest, manage and preserve digital objects, and for a Designated Community to access them and understand them. OAIS

⁶⁸ Dale and Ambacher, "Trustworthy Repositories Audit," 6-7

⁶⁹ Bak, "Trusted by Whom," 397-8

⁷⁰ Luciana Duranti, "Reliability and Authenticity: The Concepts and Their Implications," *Archivaria* 39 (Spring 1995): 7-8.

provides an approach that is open ended but directed enough to give digital repositories a way to pursue digital archiving, while at the same time giving them the freedom to work within their means.

These standards are widely available, widely used and recognized by archivists, and combining sound theory with sufficient openness to make them more generally applicable. They provide the freedom for institutions with many or few financial and other resources to actively engage in and pursue digital preservation in a way that is possible for small and large repositories alike.

Available Digital Solutions:

As digital preservation and the practices and theory around it evolved two approaches to this issue developed. First there are those that embraced the thinking behind the National Archives of Canada, Pittsburgh and University of British Columbia projects. This was a view that advocated for overarching digital systems that would help create and control digital environments and allow a more automated approach to digital archiving. The second way of thinking emphasized more local and small scale solutions, based on practicality and open source software, as the best way to push digital preservation forward. This approach to digital preservation was intended to be cheaper and less resource intensive for smaller institutions. I will first explore the more established and automated software suites that have become popular with some Canadian archivists before exploring the second approach that, as in this thesis, advocates for a more hands on approach. The purpose of exploring these solutions is to give an idea of what the current landscape is for digital solutions. Exploring in some detail the nature of

these solutions shows how the archival community has been trying to solve the issue of digital preservation and puts DIY in a more contextualized position where its strengths are more evident. This section will showcase the content based approach that some are taking with systems like Preservica and Archivematica, and the more supplemental approach to strengthen current programs that BitCurator offers.

In 2009, Artefactual Systems released the digital preservation software suite Archivematica. Their design for this system was to combine several open source digital preservation tools to create a software suite for digital preservation. Rather than having to run command line prompts, create metadata, XML files, and other somewhat difficult computer processes Archivematica allowed for users to direct the software suite to the file they wanted to preserve and then, using automated functions, to process it. The idea was to provide a graphical interface that would allow even those that lack advanced computer skills to process digital records for preservation and access.⁷¹ The plan to develop this software suite originally came in response to a 2007 UNESCO report on open source digital preservation titled *Towards an Open Source Repository and Preservation System*. This report analyzed a variety of digital tools available in 2007 and concluded that the current tools were not comprehensive or integrated enough to provide a complete digital environment that could handle the processes necessary for digital preservation from ingest to

⁷¹ "Archivematica." Accessed December 16, 2017.
https://wiki.archivematica.org/Main_Page.

preservation.⁷² In response Artefactual Systems worked with a variety of backers and co-contributors to build up Archivemata as an open source solution. It was designed to handle digital objects from ingest to access, streamlining and simplifying an open source method of digital preservation.

Archivemata runs on the Ubuntu Linux computer operating system. For it to function it has to be run on this operating system or in an application like Virtual Box that allows for an artificial environment on the computer where one can run programs from a non-native operating system. This proves to be a hindrance to the software suite since many institutions, due to corporate cultures or security concerns, prefer to run Apple or Microsoft operating systems. This can make it difficult for some to get official approval to run a computer with Ubuntu operating systems and use the software, even though many of the programs used for file migration are available on Apple and Microsoft operating systems. After one installs Archivemata from the Artefactual site through a series of command line instructions that orders the computer to find and install the suite it becomes available to use through an in-browser hub. In this hub users set parameters like storage locations, formats for access and preservation, and can tweak what metadata they will produce. To process a file through Archivemata one would click under the transfer tab in the browser and select the file they want to preserve. Archivemata will analyze the file and create what in OAIS is called a Submission Information Package (SIP). This is then selected under the ingest tab of

⁷² Peter van Garderen, "Archivemata: Using Micro-Services and Open-Source Software to Deliver a Comprehensive Digital Curation Solution." *Proceedings of the 7th International Conference on Preservation of Digital Objects*, 145–49. Austrian Computer Society, 2010, 145-7.

Archivematica, where it is further processed by migrating file formats, creating metadata, and ultimately placing all the data (originals, metadata, and migrated files for access and preservation) in .tar.gz, an archived or compressed file format used on the Ubuntu OS.⁷³ Using these compressed files an archive is able to store preservation and access copies of their digital records.

A preservation format is usually called a lossless format, such as a .wav audio file. Data is less compressed in this format meaning more of the original data is preserved. For example most CDs come with the audio files formatted into the lossless format FLAC. This allows for the audio files to be of the same quality as their original source. These lossless formats tend to be larger in file size and may require specialized software to access, making them more difficult to work with.⁷⁴ An access format is one that is considered widely accessible on modern computers, making it a kind of display format that the vast majority of people can either access through basic programs or easily find a program to install to access the file. These are usually lossy formats as opposed to lossless. These are referred to as such because they compress data sets with algorithms that result in smaller files and some lost data. An example of this is the audio access standard.mp4. These are easily accessible

⁷³ "Archivematica Documentation | Documentation (Archivematica 1.6) | Archivematica: Open-Source Digital Preservation System." Accessed December 16, 2017. <https://www.archivematica.org/en/docs/archivematica-1.6/>.

⁷⁴ Chris Hoffman, "What Lossless File Formats Are & Why You Shouldn't Convert Lossy to Lossless." Accessed December 16, 2017. <https://www.howtogeek.com/142174/what-lossless-file-formats-are-why-you-shouldnt-convert-lossy-to-lossless/>.

through most free audio players. This is the standard format of audio files on iTunes.⁷⁵

Archivematica has become a popular program because it seems to address some of the issues that were raised by big digital archiving projects of the 1990s. It creates an automated system that in theory allows for the ingest, migration, and storage that is considered the key in preserving digital records. Features other than the automation of this process have made it popular as well. Artefactual describes Archivematica as a standards-based and open source tool set that is compatible with hundreds of formats and third party software. This project has the backing of some big names in the digital archives community including the University of British Columbia, Yale University, the New York Public Library and UNESCO.⁷⁶ It is compliant with key archival standards including OAIS, PREMIS, Dublin Core for descriptive metadata and other highly respected and widely implemented standards. As an open source product, not only is Archivematica built on open source tools, it is also released under GNU Affero General Public License.⁷⁷ This means that Archivematica users are not only able, but are actively encouraged to study and improve the code of the suite. Artefactual encourages users to submit possible additions to the code, making it a truly communal piece of software for archivists.⁷⁸ The Artefactual team has been able to build into Archivematica

⁷⁵ "Format Policies - Archivematica." Accessed December 16, 2017.
https://wiki.archivematica.org/Format_policies.

⁷⁶ "Archivematica." Accessed December 16, 2017.
https://wiki.archivematica.org/Main_Page.

⁷⁷ Ibid

⁷⁸ "Contributors - Archivematica." Accessed December 16, 2017.
<https://wiki.archivematica.org/Contributors>.

functionality to be compatible with literally hundreds of digital file formats. This is an important part of why Archivemata is useful. Rather than having to seek out a new tool for every migration from one format to another, Archivemata provides the tools to process most file and format types in one suite, which some see as a huge advantage. Finally, the team at Artefactual has done a good job of integrating Archivemata with other pieces of software. Archival access system AtoM, another Artefactual Systems product, uses metadata produced by Archivemata to create a web based catalogue of archival holdings and to distribute access copies to users. Another example is their integration with DSpace, a non-profit, open source repository software package that provides the tools to build digital repositories.⁷⁹

Despite these positives there have also been challenges when adopting Archivemata. For example, the Preserving (Digital) Objects With Restricted Resources (POWRR) project tried to implement Archivemata as a free solution for under-resourced heritage institutions. While they appreciated the many standards that the suite adheres to, they had technical difficulties running the suite.⁸⁰ Some tried to run it in a virtual machine environment and said that without local technical support or paid support from Artefactual, they could not have run the suite. Their final recommendation saw three of the six testing institutions conclude that Archivemata was straightforward, relatively easy to use, and found it had attractive features. The other three institutions felt it was difficult to understand

⁷⁹ "Archivemata." Accessed December 16, 2017.
https://wiki.archivemata.org/Main_Page.

⁸⁰ Jaime Schumacher, Lynne Thomas, and Drew VandeCreek. "From Theory to Action: 'Good Enough' Digital Preservation Solutions for Under-Resource Cultural Heritage Institutions." Illinois: Institute of Museum and Library Services and POWRR, 2014, 14-15.

and operate, with one specifically pointing out that it required a knowledge of command line interfaces, which they worried the average archivist or librarian may be not be able to use.⁸¹

Members of the Council of Prairie and Pacific University Libraries (COPPUL) also evaluated Archivemata and reported three principal problems. The first was that one institution had trouble with their legal team in assuring them that Archivemata met and was consistent with the privacy policies of their institution. Second, was that implementing Archivemata requires more than technological infrastructure. Without proper digital preservation policies or frameworks in place at an institutional level the Archivemata tool set makes little difference. It is true enough that there is more required than simply having a technological infrastructure for digital preservation, but Archivemata advertises itself as a total digital solution. Finally, COPPUL found that integrating Archivemata with digital repository software such as DSpace was difficult and required working closely with institutional technical support.⁸²

On a more user-based level, in his masters thesis, Chris Zaste describes a case study of his experience installing and using Archivemata at the Mennonite Heritage Centre Archives. He cited several issues, including problems with the installation of Archivemata (though this was using a virtual environment), issues

⁸¹ Ibid, 7-9.

⁸² Bronwen Sprout and Mark Jordan. "Archivemata As a Service: COPPUL's Shared Digital Preservation Platform." *Canadian Journal of Information and Library Science* 39, 2 (June 2, 2015), 242-3,

processing videos and larger files, and dramatic increase in end file size.⁸³ Zaste's experience and these studies demonstrate that digital archiving cannot be solved by simply having a tool set. It requires institutional policies and knowledgeable archivists that can work with the tools.

While Archivemata is a free and open solution, subscription-based plans for hosting services, technical support and storage options, can be attractive in the face of some of the technical issues that COPPUL and POWRR report. COPPUL, the University of British Columbia (UBC) and Artefactual came together in a partnership to provide Archivemata-as-a-service for COPPUL members. COPPUL is responsible for promoting the program, signing up institutions and providing seed money for the one time set up costs. UBC provides fee-based server hosting and digital object storage, as well as data back-up services. Artefactual provides installation of Archivemata, technical support and ongoing training for staff. This allows COPPUL members to access an existing digital preservation platform and community and technical support at much cheaper rates than hiring or creating their own systems and supports.⁸⁴

Another digital tool that archivists have actively embraced is the BitCurator software suite. BitCurator arose out of the need for the library and archival communities to have digital forensic tools with interfaces, documentation and

⁸³ Chris Zaste, "Another Bit Bytes the Dust: The Technological and Human Challenges of Digital Preservation" (University of Manitoba/University of Winnipeg, 2016), 76-80.

⁸⁴ "Archivemata," Council of Prairie and Pacific Libraries, last modified April 10, 2017, <https://coppul.ca/archivemata>. At the national level the Canadian Council of Archives also offers access to Archivemata through a tiered service called ARCHIVESCANADA Digital Preservation Service. "Digital Preservation," accessed May 18, 2018, <http://www.archivescanada.ca/ACDPS>.

functionalities that could support the workflows necessary for collecting institutions to ensure the authenticity, integrity and the provenance of digital records entering these institutions. Digital forensics focuses on the discovery, recovery, and validation of data from computer systems. This type of information would not always be immediately available to users, but would allow for the recovery, display, and recording of certain analytics about data that could help to ensure the authenticity of digital information.⁸⁵

Cal Lee, the principal investigator of the BitCurator project, explains that BitCurator built on previous work by addressing two fundamental needs and opportunities. First, they saw an opportunity for professionals that deal with digital records through integrating digital forensic tools and methods into their workflows and collection management environments. Second, they wanted to properly mediate public access to forensically acquired data.⁸⁶ Using BitCurator, Archivists can create better metadata, demonstrate authenticity, and even recover files that were thought deleted (although this does raise ethical questions). The basic plan when starting on the BitCurator project was to provide a software suite built on open source, extensible, and mature software, focusing development on extensions, plugins, and wrappers for proven forensic tools rather than developing the tools themselves. BitCurator would also adhere to common digital forensic metadata standards and

⁸⁵ Christopher A. Lee, Alexandra Chassanoff, Kam Woods, Matthew Kirschenbaum, and Porter Olsen. "BitCurator: Tools and Techniques for Digital Forensics in Collecting Institutions." *D-Lib Magazine* 18, no. 5/6 (May 2012). <https://doi.org/10.1045/may2012-lee>. 1/6 to 2/6

⁸⁶ Lee, Chassanoff, Woods, Kirschedbaum, and Olsen, 2/6

provide crosswalks to relevant library and archival metadata schemes.⁸⁷ BitCurator was designed to allow archivists the use of forensic tools and integrate them into existing workflows to parse data, allowing them discover data hidden within computer systems.

Another digital system that has been promoted as a possible solution for archives looking to deal with digital preservation is Preservica. Preservica is both the name of a digital preservation system, and the company that sells this service and specializes in data preservation. Preservica was intended to be a corporate tool, but some have argued that it may be useful for small to mid sized archival institutions as a way for them to store digital records. The POWRR group lists 21 desired functions for candidate digital solutions; Preservica claims their product meets them all.⁸⁸ As Preservica is built on a proprietary code base, and is not open source, it would be necessary to purchase an instance of Preservica to determine the validity of these claims. In their report on digital solutions for small institutions all POWRR reviewers found Preservica to be reliable, readily customizable, and useable out of the box. They commented that they liked the interface and that customer service was good.⁸⁹

Preservica works by providing users with an application on their computer. Through this application, users can tag and upload files to a cloud storage system that is managed by Preservica. While uploading there is the option to add, both

⁸⁷ Ibid, 3/6

⁸⁸ Preservica Overview - Preservica." Accessed February 21, 2017.
<http://preservica.com/preservica-overview/>.

⁸⁹ Preservica Overview - Preservica." Accessed February 21, 2017.
<http://preservica.com/preservica-overview/>.

descriptive and preservation metadata, which is intended to help with discoverability and preservation of the files.⁹⁰ This addresses several key concerns for digital collections by allowing for the metadata necessary to properly preserve a record be added and backed up in the same place as the records. An interesting part of this in terms of provenance is that Preservica recreates the original file structure of anything put into storage, thus preserving the original order. It also tags metadata to every file in an upload, making sure that the proper metadata exists at every level. The system also has the ability to set up file migration and deletion schedules based on retention cycles. The application, and the cloud-based storage provided through Preservica, allows users to process, ingest, access, and store digital materials using an OAIS compliant workflow.⁹¹ In this Preservica is somewhat similar to COPPUL's cloud-based Archivematica-as-a-service.

Current State of Digital Solutions:

This first chapter was intended to give a broad overview of digital archiving in Canada up to the mid-2000s. The historical section provided the context of where the profession was at leading into the 2000s, and the other two sections were meant to cover some of the most widely recognized digital solutions and key standards being used and/or advocated for in Canada. Through these we can get a sense of the current state of digital archiving in Canada.

By the mid-2000s, thanks partly to previous attempted projects, the standard solution for digital preservation was to attempt to create total electronic record

⁹⁰ "Preservica: The Product Walkthrough - Preservica." Accessed March 16, 2017. <http://preservica.com/resource/preservica-the-product-walkthrough/>.

⁹¹ "Preservica: The Product Walkthrough - Preservica." Accessed March 16, 2017. <http://preservica.com/resource/preservica-the-product-walkthrough/>.

management systems. This is shown in John McDonald's "Wild Frontier" when he discussed attempts by the federal archives to create a record keeping system that would also be a digital environment to easily preserve born digital records. By the mid-2000s Library and Archives Canada once again tried to create a high-level digital preservation environment, this time a Trusted Digital Repository that was known as LAC TDR. A 2008 piece by Bak and Armstrong describes this process in some detail. LAC TDR unified archival and library functions for the acquisition, processing, preservation and dissemination of digital objects. The plan was to build upon then-current LAC policies, workflows, tools, and software and coordinate these with international standards and methodologies like TRAC and OAIS.⁹² The principal tenets of this project focused on file format migration, a system to monitor the health of files, data redundancy, open standards for descriptive, preservation, technical and rights metadata, and an information sharing network for both Canadian and international TDRs. In 2014, however, the Auditor General reported that LAC, after spending more than \$15 million, shut down the project. LAC TDR never truly went live.⁹³

There are a wide variety of different digital projects at the provincial or local levels in Canada. For example, the Archives of Ontario attempted to preserve electronic records by moving them off of 5.25 inch floppy disks so the records would be more easily accessible for use and preservation. Charles Levi, one of the archivists working on this project, wrote an article that gives a broad overview of

⁹² Greg Bak, and Pam Armstrong. "Points of Convergence: Seamless Long-Term Access to Digital Publications and Archival Records at Library and Archives Canada." *Archival Science* 8 (2008): 279-82,

⁹³ Greg Bak, "Trusted by Whom," 373-376.

the project and explores the difficulty of working with legacy media, which requires specialized hardware and software to access and preserve these records.⁹⁴ The City of Vancouver Archives also had a project focusing on digital preservation that saw them trying to ingest and process the records from the 2010 Winter Olympics hosted by the city. During these Olympic games many digital records were created that had to be preserved in the city archives, such as the hours of footage of the Olympic Torch making its way across Canada.⁹⁵ These two examples demonstrate the many digital standards, solutions and challenges in Canada. These vary by archive, as do the resources available to each archives, and so a variety of digital preservation systems and methods have been developed.

This ad hoc and individualized basis has been one way of attempting digital preservation. The counter point to this method would be a more crowd sourced way where archives are able to pool resources and skills to share the work load. This is where a toolset like Archivemata would fit in. It may seem to be one of, if not the only, solution to Canada's digital archiving challenges. It provides a free, open source and standards based solution to archiving digital records. However, as the next chapter will show, this is not the whole story. There are some drawbacks and challenges with this platform that make it difficult to use, especially for smaller institutions that do not have the technical skills or support to make it work.. That a

⁹⁴ 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): 240-2.

⁹⁵ Courtney 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 VANCO_Records AS Archives WHERE Value='true,'" *Archivaria* 72 (Fall 2011): 93-4, and 119-121.

software suite like Archivematica or BitCurator meets standards like OAIS, TRAC, and PREMIS seems impressive. However these standards were designed to be quite easy to meet, and open to a wide variety of interpretations.⁹⁶ The next two chapters in this thesis will describe attempts to implement Archivematica at the Faculty of Medicine Archives at the University of Manitoba. Instead of ultimately continuing with Archivematica, we researched key standards and, inspired by the Archivematica workflow, created our own standards-based, in-house DIY workflow for digital preservation. I believe that this is where the future of the archival profession must go. We need to continue to encourage these larger software solutions that are useful in their own way, but also encourage archivists to start on the project of digital preservation. As the PAC MRA Division learned back in the 1970s, it is hard to find that rare person skilled with computers and archival theory that can be the complete digital archivist we need. Much like they had to archivists now need to work on digital preservation, but at the same time come together and share ideas and methods. Archivists should start on smaller projects that teach the skills and knowledge necessary to do digital preservation for themselves.

⁹⁶ Chris Zaste, "Another Bit Bytes the Dust: The Technological and Human Challenges of Digital Preservation" (Masters Thesis University of Manitoba/University of Winnipeg, 2016), 1-2, 41-43, and 72-3.

Chapter 2

This entire thesis project and workflow are built on the presupposition that the Faculty of Medicine Archives requires a digital preservation system. It is a reality of modern archives, now that we are thirty years into widespread desktop computing, that they need to enter into collecting and preserving digital records. Failure to do so will see mandates not being met and records being lost. Before the mid-1980s there was niche collecting of digital records in the mainframe era, but the revolution of desktop computing and the advance of the personal computer into office life and the home has made digital record keeping inevitable.⁹⁷ The mandate of the Faculty of Medicine Archives states that it seeks to document the primary functions and activities of the Faculty of Medicine in the areas of medical education and research, and in the provision of health care in Manitoba, as well as to preserve and facilitate access to the heritage of the Faculty of Medicine through its collections.⁹⁸

With this mandate in mind, and the Designated Community it evokes, it is clear that the Faculty of Medicine Archives needs some system for digital preservation. Medical education and research require computers and may have digital components if the researcher uses a word processor, e-mail, or some other

⁹⁷ Daniel Elves, "Advocating Electronic Records: Archival and Records Management Promotion of new Approaches to Long-Term Digital Preservation," (Masters Thesis University of Manitoba/University of Winnipeg, 2012), 12-13, and 27-35.

⁹⁸ "University of Manitoba - Libraries - Mission and Mandate," accessed October 3, 2017, <http://www.umanitoba.ca/libraries/units/health/archives/quickfacts.html>.

type of digital tool to aid in their research. Further, many aspects of education have moved into the digital. Grades, course materials, registration, syllabi, and other documents related to education are produced and/or distributed largely through digital means. Thus without a digital aspect the Faculty of Medicine Archives would not be able to fulfill its mandate. Further, the former Medical Archivist, Jordan Bass, has spent considerable time and effort making the archive a digital ready and focused repository. He has done both research and work in the archive to advance its digital capacity.

Our digital preservation workflow project was heavily influenced by the position of the Faculty of Medicine Archives as a unit of the Neil John Mclean Library, itself part of the University of Manitoba Libraries network. This administrative context limited our autonomy. As a small unit in the library we were not afforded the same resources or support as larger, and more used, units within the library. This lack of technical support and dedicated technical funding encouraged us to pursue a small-scale D.I.Y solution.

Lacking technical resources, support and funding, it makes sense to pursue low-cost options like Archivematica. Finances were a major constraint throughout the project. Since we could afford to spend very little, if anything, on the project, we repurposed computers that the archive already had, used free and open source software, and designed a work-flow that worked with the technology we had. Furthermore, as I was a student working on a limited-term internship, I developed the workflow such that Bass (or other staff or interns at the archives) could pick up the pieces and continue the work after my departure. This drew me towards

software solutions that were easy to use and could work with multiple files at once. In addition to being simple to use, the workflow also needed to be able to run in the background while the small staff at the archives looked after other matters.

The biggest influence on the workflow, however, was our choice to focus specifically on data rescue. Digital files from *Honouring the Voices: 40 Years of First Nations, Metis, Inuit & Indigenous Health Research in Manitoba*, an oral history project at the University of Manitoba, were sitting on a hard drive with no plans for their preservation. Since this non-redundant and unverifiable form of storage left the files vulnerable to undetectable degradation, and given the time limits of my internship, we prioritized rescuing these files immediately, while my time could be dedicated to them. This short-term focus on saving these files was the right choice at the time. However, as this chapter will show, our choice to focus on data rescue had serious impacts on the workflow, specifically with regards to long-term preservation, and especially the creation of proper preservation metadata. Furthermore we did not focus on access to the files. Since my internship ended Bass has been able to provide access to the collections we preserved, but the ability to access files was not something we focused on when creating our workflow.

This focus of my internship on digital preservation over other records is something that is backed by the archival literature as an important endeavor. Adrian Brown's *Practical Digital Preservation*, lays out the case for digital preservation and why it matters. He argues that digital records are easier to manage and provide access to than physical records because they can be accessed remotely. At any rate, the legal mandates of archives require that they acquire and preserve more and

more born digital content.⁹⁹ Moreover, Brown notes that “Digital data requires continuous, active intervention to preserve it.”¹⁰⁰ Digital is the future, for the acquisition of new records and for the ongoing management of existing non-digital collections, including their digitization. Many organizations, both private and public, heavily rely on computers and other digital resources to generate their records. There will be more and more born digital records coming into archives. The skills to handle digital preservation properly need to be developed now. This way archivists can start the process of continuous and active intervention that is necessary for digital preservation.

Brown identifies two major threats to preserving authentic, accessible digital information. These are the loss of the data object, and the loss of the information object. The data object is something called the bitstream, the literal 1s and 0s, or bits, that make up any digital object. The information object is the data object rendered in a way that is understandable to users.¹⁰¹ For digital preservation we need the hardware and software that allows for access to the record, or the record is essentially useless as no one can interact with its data in a meaningful way. Brown identifies, then, two overarching generic threats to digital preservation. These are: technological obsolescence, and inadequate skills and resources.¹⁰² The first deals with hardware, software, and formats that become obsolete as they are no longer utilized anymore by the general public. An example of this would be the switch from

⁹⁹ Brown, Adrian. *Practical Digital Preservation: A How-to Guide for Organizations of Any Size*. London: Facet Publishing, 2013, 20-24

¹⁰⁰ Ibid, 195

¹⁰¹ Ibid, 202

¹⁰² Ibid, 203

output onto magnetic tape to hard disk drives. This created a completely new hardware situation and the tools to access the data on magnetic tape fell into obsolescence. This is why file normalization (also known as format migration) is important for digital preservation. Through file normalization, archives change the format of data objects so that they can be rendered on current hardware and software. Brown's second overarching threat to digital records, the inadequate skills and resources of many archives, is all too common among Canadian archives. This was true when Brown published his book in 2013 it remains true today.

Archivists all too often lack the skills and resources to address digital preservation. This is why I wanted to explore and explain the DIY approach that Bass and I took to digital preservation during my internship during the summer of 2016 at the University of Manitoba Faculty of Medicine Archives. I hope that other archivists might learn from my experience and be encouraged to run their own experiments in digital archiving. Since one of my hopes for this thesis is that other archivists will be able to fully understand and build upon our workflow, in this chapter I provide a detailed and occasionally technical description of our decision-making and the resulting process.

This internship was funded through a Young Canada Works and Manitoba Heritage grant that allowed the archive to hire an Archival Studies student from the University of Manitoba's Joint Master's Program in History. At the Faculty of Medicine Archives I worked under Jordan Bass, who was then the medical archivist. The focus of this internship was to process one non-digital collection from ingest to access, and to aid Bass on several digital projects that already were under way when

I started my internship in early May. Relevant to this thesis was my work on processing two digital collections for the archives.

One of my first tasks was to set up a computer as a dedicated Archivemata terminal. As noted in chapter one, Archivemata requires the Ubuntu Linux Operating System (OS) to run. There are two ways that one can install this operating environment. The first is to install and utilize a virtual digital environment on a computer. The program of choice for Archivemata is Oracle Virtual Box, which allows for the simulation of a guest operating system on Microsoft, Mac, Linux, and other OS.¹⁰³ This means that on a computer running a Windows operating system it is possible to run Ubuntu and therefore install and use Archivemata. We had a computer available making a second solution, the one that Bass tasked me with, more attractive. Instead of running a guest OS, we would take a computer, wipe it and install Ubuntu as the only OS on the machine. This would give us a machine that could run Archivemata and a host of other open source tools through Ubuntu.

By using a boot ready flash drive I was able to install Ubuntu through BIOS commands. After this the next task was to install the Archivemata software suite and to set it up with a Network Attached Storage (NAS) unit. To install Archivemata on a computer running Ubuntu it is necessary to use the command terminal of the machine. In Linux there can be an application store that will serve up some applications, but the point and objective of Linux platforms has always been open computing. There are many hubs online that host open source applications and describe the steps to download the software. On the Archivemata website

¹⁰³ "Virtual Appliance Instructions - Archivemata." Accessed October 2, 2017. https://wiki.archivemata.org/Virtual_appliance_instructions.

there is a download section that provides a series of instructions on how to download Archivemata and install it by building the software suite from the downloaded files.

There were problems with this process. It took five days to actually get Archivemata installed and functioning on our new computer. At the time version 1.4 was the newest version of the suite and the plan was to download that version on this fresh Ubuntu install. Running the commands in terminal did work, but it did not install or configure the packages properly. I could not run the software suite. Since the Medical Archives lacked the resources to pay for support, I turned to the free Google discussion group for Archivemata. Unfortunately, was not able to find the help I needed there. It took a week of troubleshooting, but in the end I simply ran the commands over and over again until the install completed to the point that Archivemata functioned. At this point, then, I finally had a functioning version of Archivemata on a dedicated computer, in addition to another older computer that was running Ubuntu and Archivemata, which Bass updated to the latest version of Archivemata. Our next step was to try to process some digital records to make sure the software suite was not only functioning, but would meet the needs of our upcoming projects.

While I had some experience with Archivemata Bass had me play around with the software suite to get reacquainted with it, and to see how it would perform processing the types of files he wanted to put through it. Bass had originally planned for two digital processing projects that I would complete while I was doing my internship. The first was processing a collection of graduate photos that had been

digitized, and the second was to process the files from the Honouring the Voices project. Processing these collections would require working with large numbers of picture files in the first case, and video, audio and text documents in the second case. Two major challenges soon presented themselves for our project. First, depending on the size of a file it could take a long time to process an ingest with Archivematica. There were times I had to leave it running for hours and ingests would still fail. It was unclear whether this was a problem with Archivematica or with file migration in general. Any transformation from one format to another can take an extremely long time. This is particularly common with video, which sometimes requires a one to one conversion – which means that every second of video requires a second of time to process it – though most video files will take around a fourth of their length to convert formats.¹⁰⁴ When converting a single file of ten minutes that is not a huge issue. But when ingesting an entire folder at once, as with Archivematica, such long conversion times increase the likelihood of errors.¹⁰⁵

The second major issue we had was storage. Both computers we were using were fairly old, with only 500 gigabyte hard drive, which is why Bass purchased a NAS with four terabytes of storage. Storage space became an issue after I had a

¹⁰⁴ “HandBrake Documentation — Video Encoding Speed.” Accessed October 2, 2017. <https://handbrake.fr/docs/en/1.0.0/technical/video-encoding-performance.html>.

¹⁰⁵ In his thesis Chris Zaste cites several issues he also had with Archivematica. These include things such as: issues with larger files, issues with processing video files, and issues with the installation. However, Zaste opted to use a virtual environment to run Ubuntu, whereas we installed the operating system. Chris Zaste, “Another Bit Bytes the Dust: The Technological and Human Challenges of Digital Preservation” (University of Manitoba/University of Winnipeg, 2016), 76-78.

series of failures in processing files beyond the ingest stage in Archivematica. Some packages were simply timing out with no real indication as to why. Eventually, I realized the computer was giving an error message that my storage was full. When Archivematica processed a video file it increased over seventy times the file's original size: a one megabyte file was coming out at over 70 megabytes. Our hard drives were already full.

Our initial solution was to get the NAS unit running. This would give us more than enough room to store and normalize the files we were trying to preserve. Unfortunately, due to how the university's network was set up, it was difficult to get the NAS to connect. Eventually I created a workaround through my personal MacBook Pro laptop (the NAS never did work with Linux based OS). Our storage issues (more or less) solved, we began to address the question of how to stop our processing from timing out. At first the problem seemed to be that we did not have enough random access memory (RAM) on the computers. RAM has a direct effect on processing power, and lower RAM systems have had issues with Archivematica in the past as it is very resource intensive to run the local servers and conversion software at the same time.¹⁰⁶ Our thought, then, was to see if we could increase the RAM of the older Ubuntu computer. As it was an older model Mac, the library's IT support told us that we could not expand or upgrade it.

Eventually, I noticed that our time out issues were happening when Archivematica was using the MySQL servers. MySQL is a piece of open source

¹⁰⁶ "Normalization for Preservation Issues - Google Groups." Accessed October 2, 2017.

<https://groups.google.com/forum/#!searchin/archivematica/mysql|sort:date/archivematica/oC18Puc-nbw/I30VIjlxCwAJ>.

software that is used to store short term processing data for Archivemata. My thought was that, if these connections were timing out, then increasing the amount of connections allowed would permit the job to complete before the time out occurred. Via the Archivemata Google Group and other resources I was able to find a way to increase the amount of server connections through terminal commands.

At first increasing the amount of connections did have a noticeable impact, and for two days I had no issues with timeouts. After that the process began to fail at the same point again with server time outs. I tried to further increase the connection limits, but it did not help this time. Artefactual staff on the Google group then told me that we should not be experiencing this error and what I had done should not have been a viable solution.¹⁰⁷ We decided that, whatever its strengths as a hosted product, we could not rely on the implementations of Archivemata that we had created and were managing in our archives, not just for video formats but also for other formats.

At this point Bass and I, frustrated with Archivemata, toyed with the idea of something new. I mentioned to Bass that I was interested in building a computer from the ground up, specifically to do this kind of digital archiving work. My idea was to look at Raspberry Pi, a cheap single-board computer used to promote computer science, and see if it could be used to make a system that would run Archivemata. Bass in turn suggested that we should work out the processes behind Archivemata, cut out the non-working elements within the suite, and run it

¹⁰⁷ "MYSQL Issues - Google Groups." Accessed October 2, 2017.
<https://groups.google.com/forum/#!searchin/archivemata/mysql|sort:date/archivemata/LeoCgXPQDYI/EssKMd1EOgAJ>.

ourselves. We decided to stop using the system and instead take it apart, use the tools ourselves and build it how we needed it to work. This was only possible because of the open source nature of Archivemata and its various component applications.

The first step was to analyze Archivemata and unpack its workflow. Since Archivemata follows an OAIS-compliant workflow, we decided that I would do more research on OAIS. While doing so, I also began to identify the processes behind Archivemata's ingest function, which includes steps to normalize files for preservation and access. Another question we tackled at this point was metadata. We wanted to know what kind of metadata Archivemata collected and what we should be collecting as the bare minimum metadata standard for our workflow. We did not want to go too in depth on metadata and use a lot of our time and resources there. We both felt my work was better spent on setting up the tools necessary to understand and establish file normalization processes instead. This would allow us to start the process of preserving at-risk digital objects as soon as possible.

Archivemata uses METS, the Metadata Encoding and Transfer Standard, to structure and store the metadata the suite generates.¹⁰⁸ METS is structured as an .xml schema, featuring a file group, structural map, and place holders or sockets for the inclusion of extension schemas that provide descriptive and administrative information. This type of schema works best with a number of additional metadata

¹⁰⁸ "Archivemata." Accessed October 2, 2017.
https://wiki.archivemata.org/Main_Page.

standards including MODS, DublinCore, NISO imaging metadata, and a few others.¹⁰⁹

Bass and I decided that we would use DublinCore standards as we found them easiest to use and most relevant to our varied content. We would use Libre Office (an open source word processor) to create our metadata and save it in an .xml file, since .xml files are easy to integrate and upload to web-based services, and are a relatively secure and stable format that aids in long term preservation. Addressing the metadata issue was my first task. Since it was an area I had little familiarity with, I conducted some focused research to make sure we did the best we could.

It was much easier for me to figure out how Archivemata ingested and normalized the files for preservation and access. Through normalization, Archivemata converts digital records into appropriate formats for preservation and access. Access formats are designed to be easily accessible across many software platforms. This allows for people to easily access and use the preserved records in some way. Preservation formats are based on lossless formats that allow for the data to be better preserved in the long term.

The basic workflow of Archivemata is quite easy to follow. The first step is use a file browser to identify the target file. Archivemata analyzes this file and performs baseline diagnostics to create a Submission Information Package (SIP). Each SIP contains, in addition to the original digital objects, a range of metadata,

¹⁰⁹ "METS: An Overview & Tutorial: Metadata Encoding and Transmission Standard (METS) Official Web Site." Accessed October 2, 2017.
<http://www.loc.gov/standards/mets/METSOverview.v2.html>.

checksums, reports and so on.¹¹⁰ Archivemata is able to add preservation, descriptive and rights metadata, much of which must be manually inputted by the archivist. This information is placed in a metadata folder in a .csv file that Archivemata is able to encode into a METS .xml file.¹¹¹ The next step is to initiate normalization. Following this, metadata, the original digital objects and the normalized files are placed into an Archival Information Package (AIP). The AIP is then compressed and stored.¹¹²

The hardest part of replicating this process was not reconstructing the workflow but finding the proper tools and learning to use them. Archivemata provides a graphical interface even though many of the programs it uses for file normalization do not. Using them on their own requires that one use text-based terminal commands rather than the visual commands that are delivered by clicking on icons in a graphical user interface. It can take some skill with computers to get Archivemata running, but it can take just as much if not more to work with some of these programs individually outside the suite. However, there tend to be more resources for using these programs than there are for Archivemata. There is also the fact that we are not bound to use any one program, and over the course of this project I experimented with many different tools for any given task. In the end, our process was similar to, but slightly different from, Archivemata. The process that

¹¹⁰ "SIP Structure - Archivemata." Accessed October 2, 2017.

https://wiki.archivemata.org/SIP_Structure.

¹¹¹ "Import Metadata | Documentation (Archivemata 1.6) | Archivemata: Open-Source Digital Preservation System." Accessed December 3, 2017.

<https://www.archivemata.org/en/docs/archivemata-1.6/user-manual/transfer/import-metadata/#import-metadata>.

¹¹² "AIP Structure - Archivemata." Accessed October 2, 2017.

https://wiki.archivemata.org/AIP_structure.

led to the creation of this workflow was much like that of the early digital archivists of the MRA. I was confronted with this issue of digital preservation, and it was very new to me. I had to come up with things as I worked, borrowing from others and inventing things myself. Much like the MRA archivists and how they had to figure things out for themselves, with the exception of the conferences on digital preservation they were able to use to give and take ideas and methods.

Our workflow from transfer to ingest to completion has seven steps. It was always meant to be a loose system and because it requires normalizing many files, one at a time, there was no need for strict adherence to how it was laid out. The only exception to this are the first two steps, which I absolutely recommend doing in order.

The first step in our workflow was to make a back up. This is as easy as copying the data to another location, on the computer, an external drive, a cloud drive, or anywhere the file will be safe and accessible. This can be surprisingly challenging when using Ubuntu as external drives have many different formats and some are not compatible with multiple OS. For example the NTFS file management format may have reading or writing issues when accessed from a Mac OS, but works fine with Windows or Linux.¹¹³ It is worth checking the format of an external drive if there is the possibility of moving between operating systems, which was an issue I had moving between Linux and Mac OS. My recommendation for back ups is to use an external hard drive as they tend to be the cheapest and safest bulk storage

¹¹³ Glenn, Walter. "What File System Should I Use for My USB Drive?" Accessed October 2, 2017. <https://www.howtogeek.com/73178/what-file-system-should-i-use-for-my-usb-drive/>.

solution. This avoids bringing potentially corrupted or infected files onto a computer and does not take up cloud or local storage¹¹⁴.

Some open source software can cause deletions or corruption of files. If backups have been created, it lessens the risk as lost or corrupted files can simply be copied back in. At this point, no files should have been opened within the ingest data set. They should be in quarantine.

The second step is to run an anti-virus scan on the target data and to create checksums. In scanning for viruses we used an open source application for Linux, Windows, and Mac OS called ClamAV. After scanning for viruses we generated a report and saved it as a .txt file in our metadata folder. This report shows this step has been taken and shows if there were any viruses on ingest, something that a user of this data may find helpful in the future as they seek to use the records and understand the original computing context in which they were generated.

After the virus check, I generated checksums for each digital object as a baseline for determining fixity. Checksums allow institutions to ensure that no one has tampered with their records and to detect media degradation such as bit rot. This an extremely important piece of metadata to ensure that files are authentic, reliable and can be verified as legitimate.¹¹⁵ For our checksums we opted to use .md5. I found a good tool set online called MD5Deep. It uses command line on

¹¹⁴ It should be noted that cloud storage is not something we explored as a possible option when thinking about storing our AIPs. The major issue with cloud storage is that it is the same as storing on a server, but the archival institution will not be in control of that server. This creates an ethical and legal concern as direct control, security and access become of even greater concern when using cloud storage.

¹¹⁵ Laura Carroll, Erika Farr, Peter Hornsby, and Ben Ranker. "A Comprehensive Approach to Born-Digital Archives." *Archivaria* 72 (Fall 2011), 71

Windows, Mac or Linux OS to batch generate checksums for an entire directory and save them to a .md5 file. Another option we explored was the Bitcurator tool suite as it offers some tools that will generate checksums one at a time. Ultimately I found MD5Deep more efficient for our purposes. At this point, I felt it important to capture the fixity of the original ingest and would generate checksums on the entirety of the ingest. These were stored in our metadata folder with the results of our virus scan.

Once the data has been backed up and the preliminary scanning is complete we could move on to the next step in our workflow. However, as I was converting files individually, rather than as batches, I found it more efficient to have multiple conversion processes running simultaneously, and to continue on other tasks during conversion. So while each step described here is necessary, one can expect to be working on multiple steps at any given time.

The third step is to check the read and write permissions of the files in the ingest. If these permissions are not open, many programs cannot access the files. Though simple to resolve, it is worth looking into at this point, as it can cause trouble later. It is more likely to be an issue with private rather than institutional records.

The fourth step is where things get more complicated, as this is the point in the process where format normalization begins. At this stage in the workflow, I knew that the data set was safely backed up, free of viruses and I had a baseline measure of its fixity. I could start to open it up and look at the file formats I would be converting, so that I would have an idea of what software I would need, and how big the final product would be. Additionally, simply knowing the formats can be helpful

for time management, since video, in particular, takes more time to convert than most other formats. Therefore I would start the video conversion first and then work on manually converting the photos and other files while the video converts. Out of every SIP, Archivemataca creates an AIP and a Dissemination Information Package (DIP). The difference between these is that an AIP contains preservation and access formats, while a DIP only has access formats. These often are two different formats, as explained in chapter one, which are used because one is good for long term preservation while the other is more widely accessible. If the ingested files are already in the formats being used as preservation or access formats, they do not have to be converted. In these cases, I was able to just copy and paste the file over rather than having to normalize it. For example, we used .mp3 as our access format for audio. If an ingest had .mp3 audio files they, could just be copied, as they were already in the access format. Since we had selected .wav as our preservation format, and as it is rare to find files in this lossless format, we still had to normalize for preservation.

Once the format of the files has been analyzed and a plan for normalization has been completed, it is time to begin normalization. Specialized software for converting files for access and preservation can be used. Format policies may vary by institution. There are often diverse lossless formats that can be considered for preservation, as well as lossy or lossless formats for access, depending on the file type and what the institution wants. Bass and I chose to use the same format policies as Archivemataca, but it may be worthwhile to look at different formats and how they affect storage size, access, and quality of finished product. Any of these

factors could contribute to changing a format standard based on the internal policies, mandate or objectives of an archive.

The fifth step is to build the access and preservation directories. This can be done while the format conversions are ongoing. I chose to use the original directory as a guide in structuring our access and preservation copies, so as to preserve the original order of the files after I normalized them. This also maintained structural relationships that are required by PREMIS. For example, an ingest may have a high level folder with three sub-folders, titled pictures, music, and video. I would create that same high level folder, but indicate in its name that it is the preservation or access copy. I would then create three sub-folders in it and place the files in the equivalent folders. This would allow for easy storage, but also maintain the key archival concept of original order. By maintaining original order, archivists hope to reflect the creator's vision and intent of the folder. There is purpose in organization and we want to reflect that when we make archival records available to users.

Step six starts by placing the normalized preservation and access files into these new folders. Alongside these new folders I also created a metadata folder. In that folder there should already be the ClamAV report that was created earlier in step two, along with the checksums. At this point I would generate checksums for the preservation and access copies as well. As generating checksums can take a while, I used this time to create a .xml file in a word processor and record our descriptive metadata. The metadata we chose to collect was based on the DublinCore standard. It has 15 unique fields that generate what has been deemed as

sufficient metadata for a collection.¹¹⁶ We chose to use seven DublinCore elements: title, creator, subject, description, contributor, date and format.

The seventh, and final step, is to finalize the AIP. At this point I would have already backed-up the data, scanned for viruses, generated checksums, normalized the files, completed the metadata, and reconstructed the file structure. The final step is to compress the folder. There are two reasons why I wanted to compress the files. The first is that it can help reduce the size of the AIP. The second reason is that it binds the collection together and makes it one cohesive, discoverable unit. This makes storage easier in the long run as the AIP is smaller and is contained in only one file that is easier for humans to find and handle.

As with access and preservation formats for individual digital objects, an institution must choose what file compression format will work best for it. Not all formats are equally accessible across operating systems, and they have different rates of compression. Archivematica uses the .tar.gz format for their AIPs. This format is only accessible on Linux; these zipped files could not be decompressed on Windows or Mac. After some research, we elected to use .7zip, .zip, or .rar as our compression format, since these are all accessible on Windows, Mac, and Linux operating systems.¹¹⁷ These formats also offered decent compression rates, which fit our needs for the smallest possible files while keeping them accessible. Once the file is compressed the AIP is complete.

Tools:

¹¹⁶ "DCMI: DCMI Metadata Terms." Accessed December 16, 2017.
<http://www.dublincore.org/documents/dcmi-terms/#section-3>.

¹¹⁷ .rar requires outside software, but programs for decompression of .7zip or .zip are included with a base install of either Windows or Mac OS.

The first tool that I used was the open source anti-virus application ClamAV.¹¹⁸ We selected ClamAV for our anti-virus program because it is what Archivematica uses, is open source and available for free, and can be used on Mac, Windows and Linux.¹¹⁹ It is easy to install, with instructions available on their website for all three operating systems. It does require text-based command line instructions, which can be difficult to get accustomed to. Most software today uses some kind of graphical interface. It can be challenging to start using text commands to do this kind of work. While it is not overly difficult to learn, some people and institutions may seek out software that uses graphical interfaces instead. In this case I was able to find a third party add-on that allowed me to use a graphical interface rather than command line, which made things very easy.¹²⁰ Using ClamAV I would scan the ingested files for viruses.¹²¹ Once it had finished it will report if it has found any viruses. This report was what I included with the metadata. I used a .txt file just because it is basic and simple to display this information.

For normalizing into our preservation and access formats, we used many distinct tools. Some of these were the tools that were deployed in the Archivematica software suite. When looking into its processes I decided that some would work for our purposes, while through exploration and experimentation we found others that

¹¹⁸ "ClamavNet." Accessed December 16, 2017. <https://www.clamav.net/>.

¹¹⁹ "ClamavNet." Accessed December 16, 2017.

<https://www.clamav.net/documents/installing-clamav>.

¹²⁰ "Clamtk." Accessed December 16, 2017. <https://dave-theunsub.github.io/clamtk/>.

¹²¹ "ClamavNet." Accessed October 2, 2017.

<https://www.clamav.net/documents/clamav-virus-database-faq>.

we preferred. This freedom to choose is a great strength of our DIY process, since we did not need these programs to function together as in Archivemata.

For audio file conversion, we followed the same format policies as Archivemata (.wav for preservation and .mp3 for access) and used Audacity, which is available for free and works on the three major operating systems, and does not use command line. Its principal weakness is that conversions must be done one at a time, as Audacity does not support batch conversion. Audacity was extremely easy and fast to use. The majority of our conversions took less than a minute to complete.

There were many options for normalizing images. Gimp is available on the three major operating systems and is free. Preview is a program included on Mac OS that will also do conversions. I ended up selecting an open source program called ImageMagick. ImageMagick is a command line tool available on all OS and was created to edit the settings and style of images, and for converting the format of images. ImageMagick allows batch conversions. Bass asked me to normalize some digitized graduation photos from the 1960s and 1970s. With ImageMagick I was able to complete what was taking days with Archivemata, in less than a single working day.

Another major file type that I spent a lot of time converting during my internship was video. The program we chose for these conversions was called Handbrake. This software is available for all operating systems as a free download from their website. Handbrake allowed me to set up many conversions and leave them running simultaneously. Since video conversion takes a long time this allowed me to work on other things while monitoring the video conversions. Starting off by

normalizing video allowed me to do other normalizations, write metadata, or work on other tasks. Handbrake has a graphical interface, which makes it fairly easy to use. One weakness of the application is that, while it accepts any file format, it only allows for output in .mp4 and .mkv. This worked for us, as these were our access and preservation formats.

For word processing files, I again found that there was a choice of software. The first option is not free, but many already have access to it, especially at an archival institution. Microsoft Word is actually very capable of doing file normalization for both preservation and access. To convert a file all I had to do was resave a document as a .pdf file, which we had identified as our preservation and access format. That is all it takes. If an institution does not have access to Word, there is a free open source option that can be used called LibreOffice. LibreOffice is an open source, free software suite available on any operating system and is meant to mirror the Microsoft office suite. The conversion method remains extremely similar.

This covers all the normalization formats we had to go through with the collections I processed at Faculty of Medicine Archives. One of the final steps I went through for this process was to generate the checksums that would allow us to confirm that our data was original and authentic. To do this, we used the program md5Deep, which is available for all operating systems. It is easy to install and use with the instructions that are available on their website.¹²² The application does require command line instructions. I also found that, depending on file size, it can

¹²² "md5deep and Hashdeep." Accessed October 2, 2017.
<http://md5deep.sourceforge.net/>.

take some time to generate a batch of checksums. I was also able to run tests to determine that md5Deep was adequate for checking fixity over time, by using md5Deep to run the .md5 file against our AIP and validate changes. Going forward, it would be possible to use md5Deep to check on fixity.

The final step of generating our AIPs was compression of the entire AIP to reduce our data storage. There are many tools out there to do this. Ubuntu comes with a GNOME-based tool called Archive Manager that can access and compress into any archive file format. GNOME is a Linux tool set that is meant to help with computer administration tasks on UNIX and Linux.¹²³ Ultimately it was not the most reliable tool, but it can work. It was more effective to use the base tools available on Mac and Windows to compress into a .zip archive format. I felt this was the best method for our purposes because it is easy to access .zip files, and there are many tools that will compress to this format. Alternately, WinRAR is free and available on all OS, has a graphical interface, that allows for compression to the .rar format. It is a little tricky to get working on Ubuntu, but is another option for compression. Ultimately at this stage there are many formats and an institution needs to understand its users and their own limitations, in terms of formats, access, compression size and computer literacy, so that it can pick a format that will work for them in terms of size and access.

Conclusions:

Building a system like this, from the ground up, has advantages and disadvantages. In this chapter I have touched on some of the advantages that can

¹²³ "GNOME System Tools." Accessed December 6, 2017.
<https://projects.gnome.org/gst/>.

come with this type of D.I.Y approach. One of, if not the, biggest, advantages to using a system like this is the cost. At the Faculty of Medicine Archives, we did not have to purchase new hardware, software, server space, or anything else that can come with setting up other digital solutions. Of course, there are hidden costs such as the salary paid to employees, technical support, physical space, and other costs non-specific to digital archiving. In our case these were minimal, as it was just Bass and I working on the project, with little to no support outside free Internet-available resources. Hidden costs do exist, but they are fairly minimal. Even solutions that are free like Archivemata lack free support and can have open and hidden costs. For example, it costs money and time to attend their seminars to better understand and utilize their product. With the software we had all coming from open source online directories it was rather easy to find support either from the creators of the software or on other platforms like YouTube. This is an advantage to using programs that have a wide user base rather than those made specifically for archival digital preservation like Archivemata. Wider adoption of open source software leads to a larger community, more support, and more resources to aid in using it. This is an advantage that I do not think can be overstated. We paid for no software, hardware or support. Our only cost was the time to set up the programs and learn to use them.

Another advantage to this system is that it can be built up as needed over time. Archivemata comes as an entire software suite, but it is rare to need every conversion tool they have right from the start. For example if an institution is just working to preserve digital pictures, they do not need to download every tool I described above. They could start with ImageMagick and some other basic tools and

learn to use them. If they go forward and need other tools for format migration they can download them at that time. Moreover, our DIY solution allows for the changing out of software as necessary. If for any reason ImageMagick does not work, it can be changed for some better tool. Archivemata users must keep ImageMagick in their toolset because Archivemata is an integrated software suite. Since we did not bundle tools together as Artefactual does with Archivemata, we can change out tools at any time. If standards change, a new piece of software is released, updates stop for a program, or any other situation arises that may make someone want to change software, our system allows for that to be done. Overall, our system requires a bit more technical knowledge than something like Archivemata, but ultimately it is easier to use and troubleshoot. There is no guessing where issues are happening because steps are done one at a time by the archivist directly. If processing fails it is easy to know exactly where it is failing.

This strength of our software choices for the DIY workflow is ultimately something that can be seen as a strength of open source software in general. Continuity of services can be a major concern for digital preservation. If a company decides that something is no longer profitable they may shut it down, or raise prices, either of which could seriously harm a digital preservation project. This is something that could happen with companies such as Artefactual Systems or Preservica. However, open source software, including open source systems originally created by Artefactual such as Archivemata, will continue to exist as long as there is a community to implement and maintain the software. Building up a DIY workflow based on community-endorsed open source software, then, will be more

immune to challenges of service continuation. Further, rather than relying on a vendor to create applications or select them for the workflow, a DIY approach lets the archives pick the tools. This allows the archives to be flexible if a particular application ceases to be updated or the community abandons it.

Even though our process of digital preservation has many strengths it also has some weaknesses. A big problem with our system is that the metadata is not complex or sophisticated in any way. We choose to simply go with DublinCore and take a minimalist approach by filling out seven of the twenty two elements in DublinCore. It is worth noting that this was an active choice by us during the creation of this workflow, and that this is an easily avoided mistake. This, along with the virus report from ClamAV and the checksums, were the only metadata we took in. When compared to other metadata standards, such as PREMIS for preservation metadata or the Rules for Archival Description for descriptive metadata, this is very basic. If I ever return to this process, metadata would be the first area that would warrant improvement, and especially preservation metadata. Without the proper metadata, preservation activities like determining software dependencies, performing fixity checks and ongoing format normalization (as standards for preservation and access change) become increasingly difficult.

Another issue for our workflow is the level of technical expertise and computer literacy needed to make it work. The process requires a lot of use of terminal to give command line instruction, and it requires sufficiently understanding the strengths and weaknesses of various formats to be able to make decisions around format policies. For these reasons it requires more technical

knowledge than Archivemata. Learning how operate command line applications can be difficult and time consuming. Archivemata requires some technical knowledge during the installation and troubleshooting process, but thanks to their pipeline design, in which the output of one process becomes input to the next, and due to their graphical interface, it does not necessarily require a lot of technical skill to operate, especially if there is available in house support or if it is hosted by Artefactual. Moreover, some archives may not have the hardware or software available to them to use Linux or open source tools. It would depend on their internal IT and security policies. To use this process, then, some places may not have the administrative freedom to use the kinds of software necessary for this approach.

To conclude, our workflow is easy to follow, and with the steps and software suggestions provided here many institutions could make use of this type of process. We began by looking at Archivemata and piecing together our own workflow over a few weeks, and then improving it as we used it. If I had had more time in my internship I would have refined it further. In sharing this workflow, I hope that others can learn from what I did and make it their own. I see this as an individualistic process that can be tailored to meet the needs of any organization or person using it to tackle the pressing issue of digital preservation. That is the ultimate strength of this workflow.

Chapter 3

In this chapter I will analyze the digital archiving workflow that I described in the second chapter with reference to the three international standards that I discussed in the first chapter: OAIS, PREMIS, and TRAC. These three standards have become foundational for contemporary digital archiving. As such it is important to understand whether the Faculty of Medicine Archives workflow complies with them. This entails reviewing what each standard establishes as the necessary minimum to meet their criteria, and once defined, reviewing the workflow described in chapter two to see if it is in compliance.

OAIS:

In the first chapter, I gave an overview of OAIS, including a description of its conformance requirements. OAIS has two rather simple requirements, listed in section 1.4 Conformance, which must be met. These are to support the model of information described in section 2.2 and to fulfill the responsibilities identified in section 3.1 of the OAIS Magenta Book.¹²⁴ An archive can provide more in the way of services, but these requirements are the base level for OAIS compliance. OAIS compliance does not specify computing platforms, command languages, interfaces, storage media or formats and so on.¹²⁵ Essentially the developers of this standard wanted to maintain an open field, and to allow for a wide variety of ways to attain compliance. Section 2.2 has three subsections that explain how an institution must define key concepts around their data to preserve it in compliance with OAIS.

¹²⁴ CCSDS Secretariat, *Reference Model for an Open Archival Information System (OAIS)*, Washington, D.C.: CCSDS, 2012, 1-3

¹²⁵ *Ibid*, 1-3 to 1-4

Section 3.1 lays out six mandatory responsibilities that any institution wanting to be OAIS compliant has to reach.

As described in chapter one, the first subsection of 2.2 deals with how information itself is defined. OAIS defines information as any type of knowledge that can be exchanged, and which can be expressed as some type of data. An archive needs to provide two things for users and account for a third. The first is the Data Object, or record. The second is the Representation Information, that will allow users to render the Data Object so that the user can understand it. Additionally, the institution must account for the Knowledge Base of its Designated Community so they can provide the necessary Representation Information. The Designated Community is an identified group of potential users (called Consumers in OAIS), who should be able to understand and access the information preserved in the archive. The Designated Community can be composed of multiple user communities and is defined by the archive and may change over time.¹²⁶ An archive needs to know their Designated Community, and have some idea as to what knowledge they reasonably should have. For example if an archive acquires information that is not in a language commonly used in their community, they should have some resources to help translate it. OAIS uses the example of providing dictionaries as possible Representation Information.¹²⁷ By combining the Data Object and Representation Information an archive will create what is called an Information Object. An Information Object is the Data Object as rendered or interpreted by the Representation Information.

¹²⁶ CCSDS Secretariat, *Reference Model*, 1-11

¹²⁷ *Ibid*, 2-3 to 2-5

The second part of section 2.2 sets out rules for how to store and present data in Information Packages. An Information Package has to hold two types of information: Content Information and Preservation Description Information (PDI). The Information Package must have Packaging Information, which is used to identify the Content Information (the target of preservation efforts) and Package Description Information. Content Information has to be defined before the PDI can be created. PDI is divided into five categories: provenance, context, reference, fixity and access rights.¹²⁸ The Designated Community must be clearly identified for proper preservation efforts as this allows for the minimum Representation Information to meaningful interact with Information Objects. Representation Information and Information Objects together make the Content Information, which is packaged with PDI and made discoverable through Descriptive Information.

The third section of 2.2 details how to properly package the information from the first two sections of 2.2. Information Packages submitted to or disseminated from an OAIS need to be distinguishable. This is in the form of the SIP when submitted and the DIP when disseminated. An archive will negotiate what information they need for preservation and this will be sent to them in the form of the SIP, which will contain Content Information and PDI. One or more SIPs can then be put through a preservation process to become an AIP containing a complete set of PDI. Finally one or more of these AIPs will be converted to DIPs when they are disseminated to users. The DIP will include Packaging Information to delineate

¹²⁸ CCSDS Secretariat, *Reference Model*, 2-3 to 2-7

between requested information, and PDI to help identify and establish the context of the Content Information.¹²⁹

In regards to section 2.2 I would like to address the three sections in order as I explore whether our method of digital preservation complies with OAIS. The first part asks whether AIPs created by an institution have all of the parts necessary to preserve records long term. The Data Objects themselves are fairly easy to understand and account for. There were records that we ingested and processed through our system for digital preservation. We meet that requirement easily. The second and third requirements are where things become more complicated. We never really defined a Designated Community for our project. Based on the Faculty of Medicine Archive's mission statement I would say that we were attempting to preserve and allow for access to documents related to the primary functions of the Faculty of Medicine for heritage and research purposes, largely with a focus on academic researchers who would be using the records for some type of project or people looking for personal or larger heritage information.¹³⁰ This definition of our Designated Community, however limits us to a more specialized group, who would be familiar with using archives. This is consistent with how Designated Communities are represented in OAIS, which caters to the more specialized community that we might call primary users. Possibly due to its origins in the aerospace community, OAIS does not really allow for a broad user base.¹³¹ This is

¹²⁹ CCSDS Secretariat, *Reference Model*, 2-7 to 2-8

¹³⁰ "University of Manitoba - Libraries - Mission and Mandate." Accessed October 3, 2017. <http://www.umanitoba.ca/libraries/units/health/archives/quickfacts.html>.

¹³¹ Jerome P. McDonough, "'Knee-Deep in the Data': Practical Problems in Applying the OAIS Reference Model to the Preservation of Computer Games," In *System*

evident by considering the concept of Representation Information. In our case, the Faculty of Medicine Archives are able to meet the needs of this primary user community for Representation Information, by having tools like dictionaries and medical texts on hand, as well as the software to access our digital records. Thus we met the first part of section 2.2, as intended by OAIS.

The second part of 2.2 defines an Information Package and its contents. Each package requires Content Information, Packaging Information, Descriptive Information, and PDI. Content Information is accounted for in my discussion of the first part of 2.2; our Representation Information and records together make up the Content Information. PDI is made up of five categories: provenance, context, reference, fixity and access rights. For provenance our metadata included the category of creator. This is part of the provenance, although in some cases it would not be the entirety of the provenance. In the future then we might want to broaden that field in some ways, such as including sections to note when the record was archived, modifications within the archive, and other important events in its management, to make it more compatible with OAIS. Context describes the context of records creation and possibly relates a data set to another one outside its AIP. This is the type of information we included under the description heading in our metadata, where we described what the data was and some aspects of its creation. Again this is rather weak and, moving forward, this is something that probably deserves its own category within our metadata file. Reference is a unique identifier for the data. We were working with files created for the University of Manitoba that

Science (HICSS), 2012 45th Hawaii International Conference on, pp. 1625-1634. IEEE, 2012, 1629-30.

would not have official identifiers like ISBNs. Bass and I discussed the possibility at one point of using software to randomly generate unique unit identifiers, but we never moved forward with this. In the future then when working with this system it will be necessary to identify an open source program that randomly generates unique identifiers, and incorporate it into the workflow. Fixity allows for the detection of change in the data; the example provided is checksums. While we generated checksums within our workflow, we did not account for degradation over time. Therefore, the workflow would have to be refined in the future to have some method of periodically re-running the checksums to verify the integrity of the data. The final piece, access rights, was not something that we delved into. We were more concerned with saving these vulnerable records and getting them to a point where we could preserve them, understanding that we would then worry about access. This does present a major issue with regards to OAIS compliance, as OAIS defines long-term preservation in terms of access.¹³²

The third part of 2.2 differentiates Submission, Archival and Dissemination Information Packages (SIPs, AIPs and DIPs). We did maintain this differentiation. We took in SIPs and processed them into AIPs. Our SIP was the files we were ingesting, and the AIP was those files with attached metadata, normalized into preservation and access formats, all in a zipped file. As I mentioned above, we were not concerned about access at this point and have yet to come to a solution as to what our DIPs would be like. The working plan was to simply give access to the completed AIPs to those that requested them. This solution is not ideal, as it raises

¹³² CCSDS Secretariat, *Reference Model*, 1-8 and 3-6

questions around giving users access to everything in the AIP, which could be confusing, and creates possibilities for users to tamper with the AIP. Again, in the future I would want to put more thought into how we might disseminate information to users.

With regards to 2.2, then, there is considerable room for improvement. In particular, the five categories of PDI should be given much closer consideration and should each have their own section within our .XML metadata document. This way we can be sure that this aspect of OAIS is met. Also lacking was our consideration of access and information dissemination. With the contract for my internship being so short I am not sure how much more I could have done. Bass and I both felt that we needed to dive into the work and start preserving these records now and worry about some things later. Access fell by the wayside because of this necessity of the work. Going forward, then, much more thought and effort needs to be put into access, including how we keep track of access rights and enable information dissemination.

Section 3.1 is the simpler part of OAIS. It lays out six mandatory responsibilities that an institution must fulfill to be considered OAIS compliant. First, an institution must negotiate for and accept the necessary information from producers. What is necessary information would be defined by how an institution defines the required Representation Information and the Information Object. Second, an institution must obtain sufficient control over the information so as to provide long-term preservation. Third, the institution must identify their Designated Community and determine its Knowledge Base so they can provide

proper Representation Information. Fourth, they must ensure preserved information is independently understandable to the Designated Community. Fifth, they must follow documented policies and procedures that ensure their records are preserved against all reasonable contingencies. Finally, they must make their preserved information available to their Designated Community, ensure that the information can be disseminated, and have evidence to prove that the information is authentic.¹³³

The first part of 3.1 is the expectation that the institution will accept required information. We did this for the two collections I processed during my time at the archive by taking in Information Objects and any Representation Information we felt we needed at the time. The second part requires that an institution gets necessary control of the documents for long term preservation. We had the original files given to us. Since we were part of the same institution that created them, we had sufficient control, physically and legally, to preserve the records. In the future it will be necessary to think about how we enact legal transfers with this system, what controls we need, and how we can ensure the records are authentic. This may be addressed through a pre-ingest checklist.

The third responsibility deals with Representation Information. We felt that we had a wide potential community of users already prepared to come into the archive and use physical records. To deal with digital records we provided a workstation and the necessary programs to access the files. While this does provide some Representation Information for our user base in the future we may have to

¹³³ CCSDS Secretariat, *Reference Model*, 3-1

provide more. A better solution going forward would be to survey our users and, from the survey results, decide what Representation Information they need.

The fourth point raised in Section 3.1 is to ensure the information can be understood on its own. We felt that what we provided through our metadata and the digital objects themselves in the AIPs it would be easy to understand the information. The fifth point is to ensure that the records are protected against reasonable contingencies. To us this meant keeping backups that could replace files if they were corrupted or damaged in some way. We did this through the use of a network attached storage unit and backing up the files to multiple places. Additionally, we relied on checksums to create a way to track changes in our records, whether these changes were the results of tampering or data degradation.

The sixth and final criterion is where our system again runs into difficulties, as OAIS requires that we provide access to these files to our Designated Community. As I have stated we were simply concerned with preservation to start, and intended to look into access later. During my time at the archive I do not think we could reasonably argue then that our Designated Community would have had adequate access to these files. However, since I have finished my internship Bass has been able to provide access to these collections through a computer in the reading room at the archive.

Overall, then, we came close to being complaint with OAIS, though there is still work to be done in the area of access rights and procedures. The nature of our work made us put preservation first, and by the time I had left the archive we had preserved the files. This allowed us to save vulnerable records. Metadata to support

long-term preservation was another area of concern. In the future preservation metadata and the five requirements of the PDI will be key areas to improve.

PREMIS:

The next major standard I will consider, PREMIS, deals specifically with preservation metadata, which provides information necessary to account for and enable long-term management. As I stated in chapter two, metadata was a weak area of our process. Again, in our drive to create a system that would address the immediate risks to these records we pushed metadata to the side. Our system suffers because of this. Analyzing our process against PREMIS illuminates how we might improve our approach to preservation metadata.

PREMIS defines itself as a comprehensive and practical resource to support the implementation of preservation metadata for digital preservation schemes. This metadata is defined in four ways: first it supports the viability, renderability, understandability, authenticity, and identity of digital objects. Second, it represents the information necessary for a repository to preserve digital objects in the long term. Third the standard emphasizes implementable metadata that is strictly defined and supported by guidelines for creation, management and use with an orientation toward automation; and finally this metadata must embody technical neutrality in that no assumptions can be made about preservation, strategies, technology or other similar aspects of institutional decisions.¹³⁴

To comply with PREMIS our metadata would have to meet these four criteria. The first point tells us what metadata has to support through the

¹³⁴ PREMIS Editorial Committee. *PREMIS Data Dictionary for Preservation Metadata*. 3rd ed. Library of Congress, 2015, 1.

information it provides, namely the five requirements listed in the first part of the metadata requirements. Our workflow somewhat meets these five requirements, but has room to improve. It supports viability and authenticity because we included a virus scan and checksums with our metadata. Authenticity is the idea that a record is what it claims to be, that there is no manipulation, substitution or falsification after its creation. This is how Luciana Duranti classifies authenticity.¹³⁵ Checksums are key to maintaining authentic records because they allows archivists to determine whether records have been changed since they were ingested into an archive. Renderability was not something that came up when we created our metadata requirements. To me this seems more applicable to records that require specialized hardware or software. However, even mainstream software and hardware does become obsolete over time. In the future, then, we would have to address not just the renderability of archival objects that require specialized tools, but also keep in mind the evolution of more common software and tools, especially as this could affect our choice of access formats. Addressing these needs in the metadata is important for users to understand how they can access the data stored in an AIP. While this did not present an issue during my internship it is something we should have addressed as there are many specialized pieces of software and hardware used within the medical field that we could encounter. Therefore in the future this is something that should be included within the metadata, in its own field. To support understandability to me suggests that PREMIS requires metadata to provide proper contextual information, similar to the Representation Information

¹³⁵ Luciana Duranti, "Reliability and Authenticity: The Concepts and Their Implications," *Archivaria* 39 (Spring 1995): 7-8.

of OAIS. While we provided some of this, not all of it was available through our metadata. Going forward our system could benefit from having more contextual information in the metadata itself. Finally identity is an area we were sufficient in. We provided dates, creator names and unique identifiers for our files in the completed metadata. This helps to keep the data identifiable and distinct. Considering how PREMIS defines metadata, then, we do not entirely measure up to the definition. PREMIS compliance is challenging even for established digital archives.¹³⁶ Acquiring new metadata will be difficult and time consuming, especially retroactively. Ultimately, some minor changes to how we planned our metadata could make this possible.

PREMIS defines four entities and three relationships as important to digital preservation through metadata. The four entities are objects, events, agents and rights.¹³⁷ Objects are the discrete units of information that are themselves being preserved, such as the actual bitstreams in the AIP. We recognized them with our metadata by describing their context by naming the creator, their date of creation, and some other factors that give further information.

Events are actions that involved an object or agent known to the system. They are represented by contextual information in the metadata that documents events throughout the management of the objects.¹³⁸ Examples include the ingest date to the archive, virus scanning, checksum generation and format normalization. To properly address this I would add a category in our metadata file that tracks

¹³⁶ Devon Ray Donaldson, and Elizabeth Yakel, "Secondary adoption of technology standards: The case of PREMIS," *Archival Science* 13 (2013), 55-58.

¹³⁷ PREMIS Editorial Committee. *PREMIS Data Dictionary*, 6-7

¹³⁸ PREMIS Editorial Committee. *PREMIS Data Dictionary*, 6-7

anything that had happened to an object since its ingest into an archival institution, including things like the date it was first accessioned, the date of its virus scan, the date the checksums were created, the date format migration took place, and other modifications and actions taken by the archival institution. Documenting such events would provide a much better picture for users, and for future archivists, as well as help to round out our metadata and bring it more in line with PREMIS.

Agents include people, organizations, software and others that have relationships to the objects.¹³⁹ Agents are somewhat represented by the creator section of our metadata, though we did not make note of our own interventions with the records. When expanding beyond people to include software this becomes much more complicated and would require more in depth metadata than we had. The “previous events” section mentioned above would go some way to addressing this point. However, providing an agents section in the metadata would work best to ensure that all agents and relationships are recorded.

The final entity is rights, which includes the assertion of legal rights or permissions. PREMIS is focused upon rights in terms of modification and ownership.¹⁴⁰ These are the most important for preservation and thus belong in a preservation metadata scheme. The creators of PREMIS largely left access rights for other systems to fulfill. We were not concerned with rights since these records were from our own institution. But even with records from our own institution we should have been thinking about intellectual property and copyright, which affects not only content but also the software used to access and store the digital records. In the

¹³⁹ Ibid

¹⁴⁰ PREMIS Editorial Committee. *PREMIS Data Dictionary*, 6-7

future it might be worth having a the donor agreement or records schedule included as part of the metadata. There should also be a section in the metadata where the rights are outlined to give archivists a brief summary of what they can and cannot do in terms of preservation activities.

PREMIS defines two types of relationships that are important in collecting preservation metadata: structural and derivation relationships. Structural relationships show the relations between parts of objects and between files that constitute a representation of an intellectual entity.¹⁴¹ This would allow for the possible rebuilding of original data sets because the structure would be known. We accounted for this with our system, but it was not recorded in the metadata. When I structured our AIPs I created new folders for the access and preservation files. In these I copied the structure from the original SIP and used it to rebuild the files as they were, thus preserving the structural relationship. This structure is also represented in our metadata through the checksums. MD5Deep allows for the inclusion of file locations when creating checksums. This means that our checksums could include location, giving a secondary means of establishing this relationship. In the future, we could add a step to our process where we would use MD5Deep to create a file listing that shows the original structural relationships of a SIP at ingest.

Derivation relationships are the result of replicating or transforming an object. In these instances the intellectual content remains the same, but its format is subject to change. The reason for recording this change is that PREMIS allows for file migration, but not the changing of a file. This is because PREMIS only allows for

¹⁴¹ Ibid, 17-20

creation, not modification, dates. This means that a file can never have dates recorded as to when changes took place, only when the file was created. PREMIS makes it necessary to create a new file, then, and this creates a need to link these files through a derivation relationship.¹⁴² We somewhat did this, but again not through our metadata. Instead, when creating the AIP we copied the original structure of the SIP with our preservation and access copies. From this a Derivation relationship could be inferred. In the future I would include it with our other metadata in some way to make it a clearly stated and defined relationship.

There is also a third type of relationship included in PREMIS, called a dependency relationship. This is when an object relies on a specific piece of hardware or software to be accessed.¹⁴³ All digital objects are in some way dependent. This could be in the form of necessary hardware, peripherals, operating systems, software, or other dependencies. Our file types are widely used by popular software options. This is one reason why many of the formats we used were the same as in Archivematica. For our more short term goals, thinking about these dependency relationships was not critical. But for long term preservation, as PREMIS supports it, there needs to be thought given to these relationships. Therefore we need to include some way to identify these relationships. There should be a system in place to keep track of the dependencies of our formats. Ideally these formats will be in common use and easily accessible for a long time, but there always needs to be the assumption they may change. This should be reflected in the metadata and preservation planning of any good digital preservation plan.

¹⁴² PREMIS Editorial Committee. *PREMIS Data Dictionary*, 17-20

¹⁴³ Ibid

Our preservation metadata, then, was very weak. I have already suggested that our barebones approach created problems for access. Ironically, our focus on data rescue could also be problematic for long term preservation. We ended up gathering sufficient descriptive metadata, but could not satisfy the preservation metadata requirements in PREMIS. Without proper preservation metadata long-term preservation can become extremely difficult. If Bass, as the sole employee, left the archives, much of the information would not be available for his successor. There is also the fact that without some of this information certain preservation activities may become difficult in the long term. Ultimately this is why PREMIS is so important to archivists. It gives us the information that we need to effect long-term preservation. The end goal of archives has always been long-term access to its records, for as long as necessary. Long-term preservation of digital objects is a hard task; PREMIS is a good starting point.

TDR TRAC:

The final standard considers not just the workflow, but the archival institution itself. It is the Trustworthy Repositories Audit and Certification: Criteria and Checklist (TRAC). TRAC lays out three broad categories that outline the basic responsibilities of an institution to meet its standards and be labeled as trusted. These categories are organizational infrastructure, digital object management, and technologies, technical infrastructure and security.

Section A on organizational infrastructure goes far beyond the scope of a digital preservation system and instead deals with issues of organizational viability,

legal and financial situations.¹⁴⁴ While this may appear to raise issues outside of our workflow, in a sense it does not. Having the proper resources ready and available to the institution is extremely important for effective long-term digital preservation. Reading through section A it become very clear that these organizational issues pose a fundamental problem for how we approach our workflow. The requirements listed are reasonable, and are an important part of creating a trusted repository. At the same time, many of the expectations expressed throughout the requirements in this section are unrealistic for most archives.

There is an emphasis in the section on the archive having control over and access to stable and reliable funding. Some may argue that such expectations privilege larger institutions, like those at the federal or provincial level. But the truth is that even those archives are not necessarily secure in their funding. Baldwin cited a 9% funding cut as a major reason as to why the MRA was closed in 1986.¹⁴⁵ There were also the budget cuts in 2012 to Library and Archives Canada that amounted to about 10% of their overall budget or 9.6 million dollars. This was to be phased in over the financial year of 2014-15. This caused LAC to estimate that their spending power had been reduced by 30%, which caused them to cut Access to Information, Circulation and Reference, IT staff, and archivists working in private acquisitions and resource discovery.¹⁴⁶ While TRAC is useful for establishing policies and

¹⁴⁴ "TRAC Metrics | CRL." Accessed October 3, 2017. <https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac>.

¹⁴⁵ Betsey Baldwin, "Stepping off the Paper Trail? Rethinking the Mainframe Era at the Public Archives of Canada," (Ph. D History-Specialization in Canadian Studies), University of Ottawa, 2006), 251

¹⁴⁶ Greg Bak, "Trusted by Whom? TDRs, standards culture and the nature of trust," *Archival Science* 16 (2016):390-1.

workflows that support trust and can contribute to an institution become trusted over time, section A ignores the harsh reality of archives. The Faculty of Medicine Archive, for example, is underfunded and has only one employee who relies on grant and heritage funding to take on projects that show the value of the archive to the community to create some public support. While section A requires permanent employees with established job descriptions, Medical Archivist Jordan Bass relied on grant-funded internships to help advance digital archiving.

Archives rely on government and heritage grants, and on donations, to get their funding. This is not the kind of stable and reliable funding demanded of them in section A. Section A4.1 requires a long term business plan, which is hard to have when sufficient, stable funding is not available. There is also in section A3 the requirement that an archive have review processes for budgets, policies, procedures, and so on. While this is important for trust, without a sizeable pool of employees this may not be possible: a one-person team with periodic interns and volunteers would have difficulty maintaining a consistent review schedule.

Ultimately then, while section A is important, comparing it to the Faculty of Medicine Archives would not be worthwhile for this thesis. It would end up being a listing of requirements that either are not met, would not advance the goal of DIY digital preservation, or that I could not validate as I lack access to the necessary information. Ultimately then there is little to justify an in-depth treatment of this section.

Section B covers digital object management, addressing how an institution handles the objects that will be preserved from ingest to access, including storage.

The first section ensures that an institution properly ingests digital objects. TRAC wants institutions to clearly identify not only content, but more specifically what properties they are preserving (section B1.1), what they will need to take in as an ingest (B1.2), that these objects are authentic and complete (B1.3 and B1.4) that they obtain enough control over a digital object to preserve it, and that they can demonstrate when they have taken control and responsibility for the record (B1.5 and B1.6).¹⁴⁷ Identifying what properties will be preserved concerns the functionality and content of the records. The point here would be that we would want to preserve the content so it remains accessible, but we would also see the functionality of the record as important. An example of this would be preserving a transcript of an audio recording. This is a viable way to preserve the content of this recording, but much of its function as an audio record would be lost. Users would not be able to listen to the actual recording, which can allow for a more meaningful interaction with the record.¹⁴⁸ With regards to the second point, we never had a system of formally stating what we were looking for in our digital ingests, nor did we have a system to prove authenticity or completeness. Ours was meant to be a small experiment to test an alternative to Archivemata and as such we neglected to think about authenticity or completeness in SIPs.

In our situation completeness would not have been as large an issue. Creating our own SIPs allowed us to create the complete record at the end of our workflow. However, the more pressing side of this would be some system to

¹⁴⁷ "TRAC Metrics | CRL." Accessed October 3, 2017. <https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac>.

¹⁴⁸ Geoffrey Yeo, "Nothing is the same as something else': significant properties and notions of identity and originality," *Archival Science* 10 (2010): 96-101.

monitor and maintain completeness over time in the face of multiple format migrations and bit rot. In the future though we would have to consider what records, and accompanying metadata, we would take in to satisfy the mandate of the archive. As for legal control over the records, this was not something we considered as the records were coming from our own institution. In other ingests there would need to be an assessment of the legal terms used for physical records and either use or update these for our digital ingests. This would also be a good way to define that we have taken ownership and responsibility of the records and that they are under our control.

Sections B1.6 through B1.8 require that a repository provides appropriate responses to donors during the ingest process, that it can demonstrate when the archive takes over preservation responsibility formally, and keeps records of any actions and processes that are relevant to the preservation of a record.¹⁴⁹ We did not formalize these processes for our workflow, but in the future this could be done through donor agreements, and through routine communications with donors. I did help to process a private collection of non-digital records for which we remained in contact with the donor to address her concerns. Finally, the last point would require setting up the same kind of process that we would need to document preservation actions under PREMIS.

The second section of digital object management, B2, deals with the creation of the archival package. B2.1 and B2.2 states that an archive must ensure there is an established definition for each AIP or class of information that will be preserved,

¹⁴⁹ "TRAC Metrics | CRL." Accessed October 3, 2017. <https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac>.

and that it is adequate for long-term preservation.¹⁵⁰ We did not do this as we were dealing with limited ingests. It would be necessary then to create a formal list of definitions for use in the future.¹⁵¹ B2.3 states that an institution needs to have a description of how AIPs are constructed from SIPs.¹⁵² This is something we fulfilled with our workflow. It would be good in the future to have a more formalized document that could be made available to potential donors, in the spirit of transparency. B2.4 requires that an institution demonstrate that all submitted objects are either placed into an AIP or disposed of in a proper-recorded fashion.¹⁵³ For our project this was never part of what we did, but I could imagine in the future creating some kind of catalogue to track this.

The next requirements, B2.5 and 2.6, deals with identifying the records. It requires that an institution either preserve unique identifiers, or use some type of naming convention for records that creates visible, persistent and unique identifiers for all archived objects.¹⁵⁴ This was addressed in the workflow through preserving the file names and structures of the ingest through to the finalization of the AIP. We used the original order of the ingest, and rebuilt that for our preservation and access file directories. This preserves unique identifiers as outlined in these sections of TRAC. B2.7 and 2.8 demand that an institution records the proper Representation Information and demonstrates that it has the tools and resources to provide

¹⁵⁰ "TRAC Metrics | CRL." Accessed October 3, 2017. <https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac>.

¹⁵¹ Ibid

¹⁵² Ibid

¹⁵³ Ibid

¹⁵⁴ Ibid

complete and authoritative Representation Information.¹⁵⁵ Our system does not meet these requirements. As discussed above with regard to OAIS we lack adequate Representation Information. Surveying our community to better understand its Knowledge Base and providing better Representation Information will also help us with OAIS as well as this part of TRAC. B2.9 further requires that there is preservation metadata for digital objects, as well as viewable documentation that shows how the records were acquired and how preservation description information will be managed.¹⁵⁶ This would be met by the changes to preservation metadata collection described above, to bring our system in line with PREMIS, as well as a publicly viewable document that explains our process for the collection of metadata. B2.10 requires a documented process for testing understandability of the information. We never considered this, but in the future there would need to be some type of process negotiated with the Designated Community for this. B2.11, 2.12 and 2.13 all deal with the AIP creation. They state that a repository must verify each AIP for completeness when it is generated, have an independent process for integrity audits, and have a contemporaneous records of actions related to preservation. We did have a process to verify completeness: at the end of the ingest process we would verify our checksums against the records and log that in the metadata. We did not have an independent mechanism for audits of our collections. With privacy legislation and other factors like financial constraints this would be a difficult to set up, though it may be possible to set up some kind of reciprocal check

¹⁵⁵ "TRAC Metrics | CRL." Accessed October 3, 2017. <https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac>.

¹⁵⁶ Ibid

process with another archive. The final point we did not cover but, as discussed in the section on PREMIS, improving our preservation metadata we would bring us in line with this requirement.

The third section, B3, of digital object management lays out some rules around preservation planning. B3.1 requires that an institution have documented preservation strategies.¹⁵⁷ Our workflow fulfills this as it is a documented preservation strategy, however, it requires an update and should be formalized. The next requirement, B3.2, is to make sure that an institution has some mechanisms in place to monitor and update Representation Information and formats as they approach obsolescence.¹⁵⁸ TRAC further requires that an institution has mechanisms in place to change its preservation plans as metadata and formats become out of date or obsolete.¹⁵⁹ We did not have a plan to monitor and update metadata and formats. However, this would be fairly easy to address in the workflow. Keeping on top of format changes and keeping a master list of formats on file would make it easy to plan when there needs to be an update. Then keeping a list of where files located based on type would be the easiest way to just go through and normalize files when the format they are in becomes obsolete. B3.3 requires a mechanism to change preservation plans as a result of monitoring activities. We did not have a formal process for this, but as we put into place more monitoring to meet the TRAC requirements it would not be hard to set up this kind of process.

¹⁵⁷ "TRAC Metrics | CRL." Accessed October 3, 2017. <https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac>.

¹⁵⁸ Ibid

¹⁵⁹ Ibid

The final part, B3.4, requires an institution provide evidence of the effectiveness of its preservation planning.¹⁶⁰ This is evidenced in our case by the collections we did preserve, and the workflows we would make public about our standards, process of preservation, and metadata schema.

The fourth section of TRAC, B4, on digital object management further addresses preservation planning. The first part, B4.1, requires that an institution employ documented preservation strategies.¹⁶¹ We implemented this in two forms: first our own documented workflow for digital preservation, and second our reliance upon recognized international standards of OAIS, PREMIS and TRAC. B4.2 requires implementing and monitoring strategies for archival object storage and migration.¹⁶² We accomplished part of this in by saving our AIPs on a network attached storage unit, with a backup copy on a second hard drive. The next requirements, B4.3 and B4.4, ensure that the institution preserves archival objects and actively monitors the integrity, or fixity, of archival objects.¹⁶³ As per our workflow we preserve content information in our AIPs. However, we had no formal plan for checking on fixity in the long term. The example given by TRAC is to maintain fixity logs, with checksums as an example. Adding a step to our workflow for the validating checksum values at regular intervals would satisfy this requirement. Finally, B4.5 requires an institution create a record of actions and the

¹⁶⁰ "TRAC Metrics | CRL." Accessed October 3, 2017. <https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac>.

¹⁶¹ *Ibid*

¹⁶² *Ibid*

¹⁶³ *Ibid*

administration processes relevant to preservation.¹⁶⁴ This aligns with PREMIS requirements of recording preservation events in the preservation metadata. This is something we did not do, but would likely try to include in the future.

Section B5 of TRAC deals with information management. B5.1 requires minimum metadata to enable Designated Communities to find material of interest and B5.2 requires that this metadata is associated with the object.¹⁶⁵ Our barebones approach to descriptive metadata meets this standard for discovery. This metadata is associated with the objects by including the metadata in our AIP. The next two points, B5.3 and B5.4, address logs and records of referential integrity for archived objects and descriptive information. They require that the institution record and provide evidence that they are doing the necessary work to keep metadata and archived objects preserved. These preservation events would include things like monitoring integrity, checking identifiers, and other activities necessary for long-term preservation.¹⁶⁶ On this point we again miss the mark. Going forward this should be a high priority task. The last part of this section, B5.6, addresses access management. As noted above we chose to focus on crisis management and short term preservation first. Access should be a priority to review in the future, and there should be consideration given to this in how we process collections, so they can be discovered and disseminated.

The third and final section of TRAC, section C, deals with technologies, technical infrastructure, and security. The first subsection, C1, focuses on system

¹⁶⁴ "TRAC Metrics | CRL." Accessed October 3, 2017. <https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac>.

¹⁶⁵ Ibid

¹⁶⁶ Ibid

infrastructure. C1.1 addresses the software that supports the institution's functions, requiring well-supported operating systems and core software.¹⁶⁷ We satisfied these criteria as we used the three major operating systems (Apple, Linux and Windows) to support our digital preservation project. C1.1 further requires strong community support as a key factor for trusting software. The commonly used open source software we used met this requirement, as there is a large community that uses this software and provides support online. C1.2 shifts to backups, stating an institution has to ensure that: it has the proper backup functionality, there has to be proper documentation of a backup policy, audits of backups and recovery plans.¹⁶⁸ We did not do this. Our backup policy was to create another copy of the final AIP to be stored on a different hard drive. To meet TRAC's standards we would have to improve on this. Related to this is the next requirement of C1.5, making sure that there is a way to detect bit corruption or loss. Data corruption or loss must be detected and reported to administration, with steps taken to repair or replace it.¹⁶⁹ Though we never planned for this, the best way to account for it would be some scheduling for fixity checks to detect these issues and in the case of loss include a formal system to report these incidents. The final points of this section, C1.3 to C1.10, deal with management of risk around the files that make up the AIP. TRAC requires that there be a plan in place for changes to the systems and files that make up the system. This plan has to continually update file formats and respond to changes in system specific hardware and software that allows for long-term

¹⁶⁷ "TRAC Metrics | CRL." Accessed October 3, 2017. <https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac>.

¹⁶⁸ Ibid

¹⁶⁹ Ibid

preservation.¹⁷⁰ Again we did not plan this far into the future. Our plan for digital preservation should include a requirement to review and update file formats. This would necessitate thinking about the financial situation of the institution and what may be possibly coming in as donations. We did not have as much financial support as many other large institutions, and that is something we would have to keep in mind when thinking about formats, hardware and software for future use. The financial situation would have to be kept in mind when purchasing hardware and software required for long-term preservation, and being able to get what is most necessary first. There is also the consideration of what will be donated, the formats these donations may take, and the hardware necessary to access and preserve these donations. Both financial costs and the usefulness of products needs to be considered in this type of plan.

Section C2 addresses appropriate technologies. C2.1 addresses hardware and how it changes. TRAC expects institutions to monitor and keep up to date with hardware used by their Designated Community.¹⁷¹ This falls outside the scope of our preservation workflow. To address it would require the Faculty of Medicine Archives to do research and acquire technology despite its limited resources. There is the possibility that records can be viewed and used without specialized hardware and software, but there could be potential issues with proprietary technologies. Therefore, research into the user base, the hardware and software they use, and how to best preserve the records would be essential in the future. C2.2 addresses

¹⁷⁰ "TRAC Metrics | CRL." Accessed October 3, 2017. <https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac>.

¹⁷¹ Ibid.

software. Archives have to ensure that they have appropriate software to provide services for their Designated Community. This would require that an archive monitors changes in hardware and software, so that an institution remains up to date and can continue to provide services.¹⁷² This is different from the hardware side of this issue because this is focused on service. As we focused on using mass market consumer software and hardware we would meet this requirement as we would have the necessary software to serve our Designated Community.

C3 deals with security. C3.1 requires an institution maintain systematic data analysis of factors like data, systems, personnel, physical plant and security needs. This would mean maintaining records of certification of key systems, risk and analysis and other factors.¹⁷³ We did not consider this when creating our system. Part of this would exist outside a digital workflow, but other areas like security and systems would have a place in it. We likely should include risk analysis as part of the workflow by having a step to analyze risk to a record and recording that somewhere, possibly in the metadata or its own location. This would be useful to aid in long term preservation, rounding out our metadata and making us more TRAC and PREMIS compliant. There is also the possibility of working with technical support within the institution if available. We did not chose to do this, but it might be possible for some archives to work with IT support, as this would fall under the responsibilities of many IT services to help maintain the physical and computer infrastructure and security needs required for TRAC.

¹⁷² "TRAC Metrics | CRL." Accessed October 3, 2017. <https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac>.

¹⁷³ Ibid

C3.2 requires an institution to implement controls that address each of its defined security needs. These include system control lists, risk, threat or control analyses, ongoing risk detection and assessment.¹⁷⁴ These were not areas we thought to address within our workflow and would have to be part of institutional-level changes. C3.3 requires that staff have delineated roles, responsibilities and authorizations related to changing the system.¹⁷⁵ Again I see this as largely a function of institutional-level systems that would not be covered by a digital preservation workflow. Areas like job roles and responsibilities seem like more an administrative area of control, however, within our workflow there could possibly be some understanding of how staff should respond to certain situations, voice their concerns, or look at changes to the system. This would require a dialogue with IT and their role within digital preservation. They would have to be informed as to what the archive was doing, their needs and the security concerns so that they can properly support digital preservation projects. Again with my work at the Faculty of Medicine Archives we did not really work with our support for this, but to better establish a permanent digital preservation plan this would have to be included in the future.

Section C3.4 requires that an institution have a written disaster preparedness and recovery plan. This plan has to include at least one off-site backup for preserved information and a copy of the recovery plan.¹⁷⁶ Part of this was

¹⁷⁴ "TRAC Metrics | CRL." Accessed October 3, 2017. <https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac>.

¹⁷⁵ Ibid

¹⁷⁶ "TRAC Metrics | CRL." Accessed October 3, 2017. <https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac>.

covered by our workflow. We would create a redundant copy of our completed AIPs for storage on a different hard drive. A recovery plan would have to have its roots and preparation in the workflow. In the future then there would have to be a formal recovery plan in place, and it would have to be integrated with the workflow. There is an official disaster policy from IST, however, with our local storage of data we would not be entirely covered by this. However, in the future to better protect and provide access to the data long term it would be useful to liaison with IST and set up a disaster plan for the archive's data that is in line with their plan.

Ultimately then our TRAC compliance would be the area in which we require the most work. That said, a lot of what TRAC targets goes beyond our workflow into legal, administrative, budgetary and other areas. This creates an issue with TRAC compliance in that I cannot entirely expect the workflow to meet its requirements. As such I would like to see the workflow meet the requirements where it can. Doing this would create a more trustworthy workflow that is better for long-term digital preservation.

This review of TRAC demonstrates that it does make a good audit tool. However, to be compliant with TRAC would require a lot of planning, preparing, and resources. To date the Center for Research Libraries has only certified six repositories under TRAC in all of North America, all of which are large repositories with substantial budgets.¹⁷⁷ Small institutions generally lack financial resources and cannot, and should not, expect to be certified compliant with TRAC. Self-audit under

¹⁷⁷ "Certification and Assessment of Digital Repositories | CRL." Accessed December 16, 2017. <https://www.crl.edu/archiving-preservation/digital-archives/certification-assessment>.

TRAC, however, seems more worthwhile for smaller institutions. For the Faculty of Medicine Archives then I would not expect our workflow to be TRAC certified, but it is helpful to use TRAC to determine gaps or weaknesses in our workflow.

Conclusion:

Our digital workflow satisfies these international standards to varying degrees. We comply most with OAIS and least with TRAC. Even though the workflow adheres quite closely to OAIS there is still some room for improvement. First, we have issues with our PDI.¹⁷⁸ Currently our metadata just records the creator, but not a lot of the necessary contextual information. We include information that would satisfy provenance because we provided the original file structure, showing how these records were organized when they came to us. We gave no consideration to access and legal rights with our data, partly because our intention was focused on data rescue, rather than longer term management and access. In the future there should be further attention paid to these rights and other aspects of the metadata that OAIS has identified as critical to long term preservation and access. With these changes to the metadata collected, our PDI would be more complete and in line with what is expected by OAIS to be collected for digital preservation. The second major issue we had was with our Dissemination Information Packages (DIPs). Since we did not consider access, we did not envision what our DIPs would look like. To be OAIS compliant access must be addressed. We need to add steps to create DIPs into the workflow. The final problem we had with OAIS was with the definitions of responsibility in 3.1. To be more compliant with this section we need to set a bare

¹⁷⁸ The five areas of PDI are provenance, context, reference, fixity, and access rights. CCSDS Secretariat, *Reference Model*, 2-3 to 2-7.

minimum of required information in our SIPs, taking into account preservation metadata requirements. There also needs to be more consideration in our plan given to legal rights that help us establish the necessary control for preservation of digital objects. Finally we need to put more work into identifying our Designated Community and determining its Knowledge Base, and arranging for their access to our collections

Our system was not very compliant with PREMIS. Since we were focused on crisis management we did not put as much thought as we could have into long-term preservation. Our immediate concern was rescuing high-risk documents and starting the process of preservation. So it is no surprise then that our preservation metadata did not measure up to the high standard of PREMIS. One of the biggest issues with our lack of preservation metadata is that we did not represent in any way the proper contextual information that is necessary for long-term preservation, addressing questions like renderability, or understandability. Including in our metadata events, agents, and rights would help provide this contextual information. Documentation of preservation events would require both a post and pre-custodial approach, to include records events from the creation of the record to its deposit in the archive and beyond. Documenting agents would require more information on the people and institutions that handled the record over its life, including in the archive. Describing rights would clarify the legal situation of the record by documenting all agreements pertaining to ownership and modifications.

Another area of PREMIS missing from our system are structural and dependency relationships. We could show the structural relationships of the SIP in

the AIP through MD5Deep. This software allows you to create a file listing that will also show the directory listing, allowing us to automate the process of documenting original order and displaying it in the metadata folder of the AIP. This would work with the simple digital objects we were processing in our pilot project, but it would not work with more complex objects like websites. To have proper metadata we should also give dependency relationships their own section within the .xml metadata sheet¹⁷⁹. Overall, then, while I do not think we were close to being compliant with PREMIS, with some small modifications to our collection and display of metadata we could better align with PREMIS. This is what makes PREMIS useful. Compliance may or may not be possible for our workflow, but by auditing it through PREMIS I have been able to highlight deficiencies in our preservation metadata. This will enable us to improve our preservation metadata, which will result in better long-term preservation.

Our workflow was least compliant with TRAC. TRAC is a tool to audit an entire institution, and not a digital archiving workflow in isolation. Some sections of TRAC are based in management, security, disaster policies, and other areas that are beyond what a workflow can include. While I do not think it is entirely relevant to the workflow, using TRAC in this way does make for some interesting insights into trust and openness. For example our lack of focus on legal rights was not an issue as the records were preserved were from our institution. TRAC highlights how important securing legal rights can be for long term preservation and trust. TRAC

¹⁷⁹ Dependency relationships are about more than hardware and software necessary to access the records. They need to be recorded to note when an Object requires an environment, or piece of hardware or software to function, be coherent, or deliver its content. PREMIS Editorial Committee. *PREMIS Data Dictionary*, 17-20.

also shows our weaknesses in metadata much like PREMIS does. It requires the collection of preservation metadata and Representation Information that would continue to help refine and improve the metadata for the workflow. Another large area that TRAC showed lacking in the workflow is continued intervention after preservation. TRAC requires a lot of monitoring, preventative maintenance, and other post-custodial activities that are key to long term preservation that are currently lacking in the workflow. Ultimately, then, to be TRAC compliant there is a lot of work to do with the workflow to make it a comprehensive alternative to other software solutions.

This analysis of the workflow against these standards is a non-traditional way of utilizing them. Many would view them as prescriptive, in that they set out a series of rules to follow to be compliant with them. By being compliant with these standards is a mark of quality within the archival community; these standards have become a way to show that a tool or solution has value and merit. However, there can be, as has been demonstrated in this chapter, a lot of value taken from simply using these standards as audit tools. This workflow is not meant to be copied one to one by anyone else. It should stand and serve as a guide of how a DIY solution can be made so others can make their own. These standards then become extremely useful audit tools, instead of marks of distinction. Since no one has to implement this process exactly this shows a way to audit against these important standards and demonstrates how that can improve a workflow. These standards have become key to digital preservation for a good reason. They provide very useful terms and ideas for which to think about digital preservation. By using them as audit tools several

flaws and drawbacks of this workflow have been shown, and that creates a way to make a better workflow. This shows the value of using these standards in this way, and demonstrates how others can do this with their DIY solutions.

Conclusion

This thesis has laid out one possible path forward for digital preservation. Rather than rely on vendors and solutions from outside the archival community, archivists can take the initiative and create their own solutions. When first confronted with computer records, MRA archivists found new and creative ways to work with the medium and preserve the records. This is something archivists today can relate to and should draw inspiration from to solve our issues. The profession should confront these records now, and try to create the relationships to share these methods as the MRA archivists did the in 1970s. Part of this can be done by building up our individual skills and resources through a DIY approach, which would further contribute to building digital preservation skills and IT literacy within the archival community. To that end it is important that the digital preservation workflow presented in chapter two, and the larger strategy of DIY in-house solutions of which it is a part, can meet the requirements of widely recognized standards for digital preservation. This is why the third chapter of this thesis may be its most important part. Auditing this workflow against three international standards, each recognized as core to digital preservation today, tests the feasibility of this type of workflow. OAIS, PREMIS, and TRAC-TDR are essential to how archives conceptualize and enact digital preservation. OAIS is important for how an archive will think about organization and packaging of their information. PREMIS creates a framework within which an archive can collect preservation metadata necessary for long-term preservation. TRAC-TDR, rather than treating the content of an archives, focuses on the institution and its policies. Together these three standards provide a total

picture of digital preservation. They represent multiple aspects that an archive has to work on if they want to do trustworthy digital preservation.

OAIS is extremely useful for auditing a preservation workflow due to its focus on the packaging, preservation and dissemination of information. Preserving information is a noble goal, but if that information is unusable or unattainable then the process of preservation is not worth the effort. The value of a record is partly determined by its possible uses, and if it cannot be used or accessed by the Designated Community then an archive is not fulfilling its role. OAIS is useful for addressing this in digital institutions in two ways. First it provides the means of packaging and receiving or disseminating information. OAIS outlines how Data Objects need to be packaged with the information required by users to understand and interpret that information so it can be useful.¹⁸⁰ There is also the distinction between how information is received, in SIPs, how it is stored, in AIPs, and how it will be made accessible and distributed to users, in DIPs. OAIS further provides a means of creating a system wherein preserved information will be understandable. This is done through the concepts of Designated Community and Representation Information. The whole idea of these concepts is that an archive has a community of users that it should be able to serve information to. Establishing a baseline of knowledge within the Designated Community and making that foundational to all services of the OAIS makes it easier for an archive to preserve information because it establishes what

¹⁸⁰ CCSDS Secretariat. *Reference Model for an Open Archival Information System (OAIS)*. Washington, D.C.: CCSDS, 2012, 2-3 to 2-8.

information is required to understand preserved content.¹⁸¹ If information cannot be understood than the preservation process has, effectively, failed.

The next standard I used as an auditing tool is PREMIS. PREMIS was designed to help institutions establish preservation metadata that supports long-term preservation. During the creation of the workflow described in chapter two we did think about metadata, but did not spend the time and effort to establish any kind of preservation metadata. This mostly came from time constraints, the urgent need to bring these records into the archives, and the fact that our project was conceived as experimental rather than permanent. I was employed as a temporary intern and, with Medical Archivist Jordan Bass being extremely busy, we recognized that if I could not process these collections Bass would likely not have the time. This is largely why we focused on such a short term and interventionist approach. But as I have stated in this thesis, this attitude and how we went about our metadata meant we could preserve these files in the short term, but set them up for failure in the long-term since, without authoritative and useful preservation metadata, long-term preservation is simply not possible. Without PREMIS metadata – including data about system dependencies and software versions – it is sometimes still possible to perform long-term preservation functions, but it becomes extremely difficult and time consuming to do so; moreover, key information about the authenticity of the records may be lost. This is why I felt it was very important to audit against PREMIS even though the workflow was likely to fail to meet its standards in almost every way. If this

¹⁸¹ Ibid.

workflow could not be made to include the necessary preservation metadata, and it was not acknowledged that there are serious shortcomings with this section in the current workflow, than it would not be worthwhile to promote this method. As it stands I think moving forward that it would take some work but that it would be possible to modify the current workflow to include preservation metadata. Ultimately, then, this shows the strength of this type of workflow. Because it is an in-house DIY solution it is a simple matter to modify and update the workflow as standards change, new ones are introduced, or an institution chooses to go with new ones.

The final standard I chose to audit against was TRAC. This audit demonstrated perhaps fundamental tensions between the standard and the workflow described in chapter two. In designing and implementing our workflow we accepted Archivematica as a base model and made its approach work better for our own use. We also wanted to make this workflow something that was cheap for us to implement, easy to use and had a DIY basis. By contrast I personally think that TRAC is an expensive and very time consuming process that focuses on the many minute, but important, details of long-term digital preservation and trust. That being said, while TRAC addresses issues of trust and digital preservation that are important to this workflow it does go far beyond this into areas of management and administration.

There is also the issue with trust itself. In “Trusted by Whom” Greg Bak explored the concept of trust at Library and Archives Canada (LAC) in their attempt to build a trusted digital repository. Even though this project was

influenced by TRAC and OAIS, there was a clear weakness noticed by Bak in that there was no clear objective in terms of trust. LAC started its project to build a TDR with the assumption that as a cultural institution with a legal mandate they would already have trust.¹⁸² As a result, the project became about the technological challenges rather than establishing a clear set of ground rules for what constitutes trust.

This same issue also is present in TRAC. TRAC is divided into three sections of requirements that are useful when building trust in a digital preservation system, addressing issues from access, funding, staffing, metadata, and other areas.¹⁸³ TRAC is useful in this sense: it would be a good place to start when building trust or a digital preservation system. However, just being able to say an institution is TRAC compliant is not a blanket statement of their trustworthiness. Taking a look at the medical archives most donors and users probably will not know what TRAC, or its administrators and auditors, the CRL, are. A repository should come at this more from the idea of the Designated Community and work with them to build their trust. This is all part of the outreach that an archive will preform.

Nonetheless, it was useful to go through an audit using TRAC. It exposed several areas of the workflow that are lacking, and provided some useful ideas of ways to improve the workflow. Undertaking this audit and using it to improve the workflow will surely lead to a better workflow that is more trustworthy. I think

¹⁸² Greg Bak, "Trusted by whom? TDRs, standards culture and the nature of trust," *Archival Science* 16 (2013): 375-79.

¹⁸³ "TRAC Metrics | CRL." Accessed October 3, 2017. <https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac>.

any small archive that does not have access to a large budget will find it is basically impossible to expect to be TRAC certified, but this does not make TRAC and its requirements useless. It has led me to make several important observations and suggest key changes to the workflow that would improve its ability to preserve digital content. However, just being TRAC compliant is not an end goal in itself. There would still need to be some kind of examination of trust between the archive and its users and donors – its Designated Community.

This idea then, of incremental improvements rather than meeting certification requirements or attaining compliance, is why I think the third chapter of this thesis is so important. Rather than using these standards in the traditional way I have taken a different approach: not to make the workflow conform with the standards, but to self-audit the workflow against the standards to identify areas for improvement. OAIS exposed weaknesses of our short term planning for access. PREMIS demonstrated how weak was our preservation metadata and provided a path to improve it so as to support long-term preservation. Finally, TRAC has shown multiple ways that are possible to improve the workflow, for example through setting up systems to check on the record integrity over time and setting up off site storage. Such measures not only improve the workflow, but can build trust in the archives. By using these standards in this way then the workflow has a path forward that will sharpen and improve it, while at the same time grant it the legitimacy of at least trying to follow the compliance of these international standards.

In 2013 Devan Ray Donaldson and Elizabeth Yakel wrote a piece on the implementation of PREMIS at several institutions as a case study for how to better adopt standards. Their point was to use case studies to point towards communal learning process that might ease implementation of these standards.¹⁸⁴ Unlike my use of PREMIS, Donaldson and Yakel were focused on actual PREMIS compliance; nevertheless I see our ideas as complementary. For example Donaldson and Yakel argue that to adopt PREMIS need not mean that an institution has to act on the entirety of the PREMIS Data Dictionary. Instead they argue for secondary adoption by using others' experiences to inform and adapt from. They claim that implementers must experiment with the standard in order to learn how to utilize it and best apply it.¹⁸⁵ I think that my audit qualifies as this type of experiment. It was a way to take the process we wanted to use for digital preservation at the Faculty of Medicine Archives and see how it measured up to PREMIS. Doing this with PREMIS and other standards allowed us to improve our practices for long-term preservation.

This conception of chapter three as an experiment into how to better utilize digital standards fits with the larger goals of my thesis, of promoting DIY, in-house solutions. Bass and I used Archivematica as an inspiration to make something that worked for our archives. We did not intend this as an exact guide for other institutions, but more like advice towards creating individualized digital

¹⁸⁴ Elizabeth Yakel, and Devan Ray Donaldson, "Secondary adoption of technological standards: The case of PREMIS," *Archival Science* 13 (2013): 55-59.

¹⁸⁵ *Ibid*, 78-79.

solutions. It is the DIY aspect that should be emulated, not our specific tools, steps, and workflow.

I believe that this is an important approach for archivists in the future. It has become inevitable that archivists will be dealing with the digital. As Laura Carroll et al. argued in their piece on the preservation of Salman Rushdie's records at Emory University, collections have moved from having a few floppy disks or CDs as "fugitive media" amidst a majority of paper files, to having complete operating systems, hard drives and entire computer systems.¹⁸⁶ If we extend this to an archive like the Archives of Manitoba or Library and Archives Canada we see a different issue. Both of these archives have legal mandates to collect provincial and federal government records. With the expansion and increasing use of computers this means that these institutions are processing digital records right now, - sometimes in the form of digital media, sometimes in the form of print-outs of digital records - and will probably continue to see the amount of born digital content increase over time. Institutional mandates, then, are requiring archives to more often work with digital records.

A DIY digital solution is something that I see as necessary to advocate for in this climate. Firstly, it puts forth a solution for smaller and under resourced archives to cheaply and effectively manage digital records. This is a way that smaller institutions will be able to continue to meet their mandates and to process digital records. Secondly, and more importantly in my mind, a solution like this leads to archivists learning digital archiving by doing digital archiving. A

¹⁸⁶ Laura Carroll, Erika Farr, Peter Hornsby, and Ben Ranker, "A Comprehensive Approach to Born-Digital Archives," *Archivaria* 72 (Fall 2011): 62-3.

DIY approach to digital preservation lets archivists slowly build up the infrastructure and skills that provide deep knowledge of how various kinds of software and hardware operate, and the skills necessary to use it.

In “Beyond the Magic to the Mechanism” Ciaran Trace states that she wants to support archivists becoming computer hobbyists who are curious to see and understand the inner workings of their computers.¹⁸⁷ Trace argues that this was normal among computer users up until the 1980s, when the Macintosh changed the relationship between computers and people. This new computing culture made it so that people only had to understand what was going on at the level of their use of digital tools, and not within the digital tools themselves.¹⁸⁸ They understand hitting a key makes a letter type, but they do not understand why. Trace suggests something that I think my workflow gets at. Archivists are going to be involved with the digital now and in increasing numbers in the future, and will require certain skills and knowledge. Trace sees the computer hobbyist as someone who knew how a computer worked and how to make it do what they wanted. Understanding how a computer works and how to work with it at this technical level is an essential skill to perform digital preservation. Trace and I both want to see archivists get more involved with computers, coding and working more in depth with hardware and software to develop the skills necessary for digital preservation. Right now, with mandates as they stand, these skills are a good thing for archivists to have. In the future they are going to be

¹⁸⁷ Ciaran B Trace, “Beyond the Magic to the Mechanism: Computers, Materiality, and What It Means for Records to Be “Born Digital,”” *Archivaria* 72 (Fall 2011): 7-10.

¹⁸⁸ *Ibid*, 6-7.

essential. That is why it is imperative that archivists immediately start the process of learning how to better utilize computers and work with the digital. DIY digital preservation offers a way to start this process of learning, and to start preserving records now.

Works Cited

- “(99+) Archivemata – Google Groups.” Accessed April 2, 2017.
<https://groups.google.com/forum/?fromgroups#!forum/archivemata>.
- “AIP Structure - Archivemata.” Accessed October 2, 2017. https://wiki.archivemata.org/AIP_structure.
- Alex, er Duryee, and Bertram Lyons. “Two Open-Source Tools for Digital Asset Metadata Management.” *Practical Technology for Archives* (blog), May 29, 2014.
https://practicaltechnologyforarchives.org/issue2_duryee_lyons/.
- Ambacher, Bruce. “Government Archives and the Digital Repository Audit Checklist.” *Journal of Digital Information* 8, no. 2 (2007).
- “Announcing 1.5.1 Release (CentOS Rpms) and 0.9 Release of Storage Service - Google Groups.” Accessed April 2, 2017. <https://groups.google.com/forum/?fromgroups#!topic/archivemata/Qkp00OnDzNw>.
- “Archivemata.” Council of Prairie and Pacific University Libraries (COPPUL), April 10, 2017.
<https://coppul.ca/archivemata>.
- “Archivemata.” Accessed April 2, 2017. https://wiki.archivemata.org/Main_Page.
- “Archivemata.” Accessed October 2, 2017. https://wiki.archivemata.org/Main_Page.
- “Archivemata.” Accessed December 16, 2017. https://wiki.archivemata.org/Main_Page.
- “Archivemata.” Accessed December 16, 2017. https://wiki.archivemata.org/Main_Page.
- “Archivemata Documentation | Documentation (Archivemata 1.6) | Archivemata: Open-Source Digital Preservation System.” Accessed December 16, 2017.
<https://www.archivemata.org/en/docs/archivemata-1.6/>.
- “Archivemata: Open-Source Digital Preservation System.” Accessed March 16, 2017.
<https://www.archivemata.org/en/>.
- “Artefactual Systems Inc.” Accessed April 2, 2017. <https://www.artefactual.com/>.
- “Artefactual Systems Inc.” Accessed April 2, 2017. <https://web-beta.archive.org/web/20030805163209/http://artefactual.com>.
- “Artefactual Systems Inc.” Accessed April 2, 2017. <https://web-beta.archive.org/web/20080621091826/http://artefactual.com>.
- “Artefactual Systems Inc.” Accessed April 2, 2017. <https://web-beta.archive.org/web/20081221075933/http://artefactual.com>.
- “Artefactual Systems Inc.” Accessed April 2, 2017. <https://web-beta.archive.org/web/20100120090650/http://artefactual.com>.
- “Artefactual Systems Inc. -- Services.” Accessed April 2, 2017. <https://web-beta.archive.org/web/20080830040748/http://artefactual.com/consulting.html>.
- “Artefactual Systems Inc. -- Services.” Accessed April 2, 2017. <https://web-beta.archive.org/web/20080830040748/http://artefactual.com/consulting.html>.
- Atherton, Jay. “Automation and the Dignity of the Archivist.” *The Canadian Archivist* 2, no. 1 (1970): 56–58.
- Baca, Murtha. “Introduction to Metadata.” InteractiveResource, July 20, 2016.
<http://www.getty.edu/publications/intrometadata>.
- Bak, Greg. “How Soon Is Now? Writings on Digital Archiving in Canada from the 1980s to 2011.” *The American Archivist* 79, no. 2 (Fall/Winter 2016): 283–319.
- — —. “Media and the Messengers: Writings on Digital Archiving in Canada from the 1960s to 1980s.” *Archivaria* 82 (2016): 55–81.
- — —. “Trusted by Whom? TDRs, Standards Culture and the Nature of Trust.” *Archival Science* 16 (2016): 373–402.
- Bak, Greg, and Pam Armstrong. “Points of Convergence: Seamless Long-Term Access to Digital Publications and Archival Records at Library and Archives Canada.” *Archival Science* 8 (2008): 279–93.
- Baldwin, Betsey. “Confronting Computers: Debates about Computers at the Public Archives of Canada during the 1960s.” *Archivaria* 62 (Fall 2006): 159–78.
- — —. “Stepping off the Paper Trail? Rethinking the Mainframe Era at the Public Archives of Canada.” Ph. D (History-Specialization in Canadian Studies), University of Ottawa, 2006.
- “Basic Usage -- IM v6 Examples.” Accessed October 2, 2017.
<http://www.imagemagick.org/Usage/basics/#mogrify>.
- Bass, Jordan. “A PIM Perspective: Leveraging Personal Information Management Research in the Archiving of Personal Digital Records.” *Archivaria* 75 (Spring 2013): 49–76.

- — —. “Getting Personal: Confronting the Challenges of Archiving Personal Records in the Digital Age.” Master of Arts, University of Manitoba/University of Winnipeg, 2012.
- Bearman, David. “Moments of Risk: Identifying Threats to Electronic Records*.” *Archivaria* 62 (Fall 2006): 15–46.
- Besek, June, Jessica Coates, Brian Fitzgerald, Wilma Mossink, William LeFurgy, Adrienne Muir, Mary Rasenberger, and Christopher Weston. “Digital Preservation and Copyright: An International Study.” *The International Journal of Digital Curation* 3, no. 2 (2008): 103–11.
- Brown, Adrian. *Practical Digital Preservation: A How-to Guid for Organizations of Any Size*. London: Facet Publishing, 2013.
- Caplan, Priscilla. “Understanding PREMIS.” Washington, D.C.: Library of Congress, February 1, 2009.
- Carrol, Laura, and Erika Farr. “A Comprehensive Approach to Born-Digital Archives,” n.d.
- Carrol, Laura, Erika Farr, Peter Hornsby, and Ben Ranker. “A Comprehensive Approach to Born-Digital Archives.” *Archivaria* 72 (Fall 2011): 61–92.
- Carroll, Michael. “The Public Archives of Canada’s Experience in Establishing a Machine Readable Archives,” 1975.
- CCSDS Secretariat. *Reference Model for an Open Archival Information System (OAIS)*. Washington, D.C.: CCSDS, 2012.
- Cerf, Vinton. “Avoiding ‘Bit Rot’: Long-Term Preservation of Digital Information.” *Proceedings of the IEEE* 99, no. 6 (June 2011): 915–16.
- “Certification and Assessment of Digital Repositories | CRL.” Accessed December 16, 2017. <https://www.crl.edu/archiving-preservation/digital-archives/certification-assessment>.
- “ClamavNet.” Accessed October 2, 2017. <https://www.clamav.net/documents/clamav-virus-database-faq>.
- “ClamavNet.” Accessed October 2, 2017. <https://www.clamav.net/documents/clamav-virus-database-faq>.
- “ClamavNet.” Accessed December 16, 2017. <https://www.clamav.net/>.
- “ClamavNet.” Accessed December 16, 2017. <https://www.clamav.net/documents/installing-clamav>.
- “Clamtk.” Accessed December 16, 2017. <https://dave-theunsub.github.io/clamtk/>.
- Cohen, Frederick. “Digital Diplomats and Forensics: Going Forward on a Global Basis.” *Records Management Journal* 25, no. 1 (2015): 21–44.
- “Contributors - Archivemata.” Accessed April 2, 2017. <https://wiki.archivemata.org/Contributors>.
- “Contributors - Archivemata.” Accessed December 16, 2017. <https://wiki.archivemata.org/Contributors>.
- Cook, Terry. “Byte-Ing Off What You Can Chew: Electronic Records Strategies for Small Archival Institutions.” *Archifacts*, 2004, 1–20.
- — —. “Easy to Byte, Harder to Chew: The Second Generation of Electronic Records Archives.” *Archivaria* 33 (Winter 1991): 202–16.
- Cothey, Viv. “Digital Curation at Gloucestershire Archives: From Ingest to Production by Way of Trusted Storage.” *Journal of the Society of Archivists* 31, no. 2 (2010): 207–28.
- Cunningham, Adrian. “Digital Curation/Digital Archiving: A View from the National Archives of Australia.” *The American Archivist* 71, no. 2 (2008): 530–43.
- — —. “Waiting for the Ghost Train: Strategies for Managing Electronic Personal Records Before It Is Too Late.” *Archival Issues* 24, no. 1 (1999): 55–64.
- Cunningham, Adrian, and Margaret Phillips. “Accountability and Accessibility: Ensuring the Evidence of E-Governance in Australia.” *Aslib Proceedings New Information Perspectives* 57, no. 4 (2005): 301–17.
- Dale, Robin, and Bruce Ambacher. “Trustworthy Repositories Audit & Certification: Criteria and Checklist.” Center for Research Libraries, 2007.
- Dappert, Angela, Sabastien Peyrard, Carol Chou, and Janet Delve. “Describing and Preserving Digital Object Environments.” *New Review of Information Networking* 18, no. 2 (2013): 106–73.
- “DCMI: DCMI Metadata Terms.” Accessed December 16, 2017. <http://www.dublincore.org/documents/dcmi-terms/#section-3>.
- Decker, A., C. Egert, A. Phelps, and J. P. McDonough. “Technical Properties of Play a Technical Analysis of Significant Properties for Video Game Preservation.” In *2012 IEEE International Games Innovation Conference*, 1–4, 2012. <https://doi.org/10.1109/IGIC.2012.6329857>.
- “Digital Preservation.” Accessed May 18, 2018. <http://www.archivescanada.ca/ACDPS>.
- Donaldson, Devan, and Paul Conway. “Implementing PREMIS: A Case Study of the Florida Digital Archive.” *Library Hi Tech* 28, no. 2 (2010): 273–89.
- Donaldson, Devan, and Elizabeth Yakel. “Secondary Adopting of Technology Standards: The Case of PREMIS.” *Archival Science* 13 (2013): 55–83.

- Douglas Engelbart | Talks at Google. Talks at Google, 2007. <https://www.youtube.com/watch?v=xQx-tuW9A4Q>.
- “Download @ ImageMagick.” Accessed October 2, 2017. <https://www.imagemagick.org/script/download.php>.
- Duranti, Luciana. “Reliability and Authenticity: The Concepts and Their Implications.” *Archivaria* 39 (1995): 5–10.
- — —. “Reliability and Authenticity: The Concepts and Their Implications.” *Archivaria* 39 (Spring 1995): 5–10.
- Edwards, Chloë, Amy F. Brown, Meg Eastwood, Martha Tenney, and Kevin O’Donnell. “Processing Internal Hard Drives.” *Practical Technology for Archives* (blog), November 17, 2013. https://practicaltechnologyforarchives.org/issue1_edwards/.
- Elves, Daniel. “Advocating Electronic Records: Archival and Records Management Promotion of New Approaches to Long-Term Digital Preservation.” Master of Arts, University of Manitoba/University of Winnipeg, 2012.
- Fiebelkorn, Guillermo. “Why Does It Take So Long? Implementing Electronic Records Programs at Universities.” Master of Arts, University of Manitoba/University of Winnipeg, 2012.
- “Format Policies - Archivemata.” Accessed December 16, 2017. https://wiki.archivemata.org/Format_policies.
- Franks, Patricia. “Government Use of Cloud-Based Long Term Digital Preservation as a Service: An Exploratory Study.” *Digital Heritage* 2 (2015): 371–74.
- Fresa, Antonella, Borje Justrell, and Claudio Prandoni. “Digital Curation and Quality Standards for Memory Institutions: PREFORMA Research Project.” *Archival Science* 15 (2015): 191–216.
- Galloway, Patricia. “Education for Digital Archiving through Studio Pedagogy, Sequential Case Studies, and Reflective Practice.” *Archivaria* 72 (2011): 169–96.
- Gilliland, Anne. “Setting the Stage.” In *Introduction to Metadata*, 3rd ed. Los Angeles: Getty Publications, 2016. <https://www.getty.edu/publications/intrometadata/setting-the-stage/>.
- Glenn, Walter. “What File System Should I Use for My USB Drive?” Accessed October 2, 2017. <https://www.howtogeek.com/73178/what-file-system-should-i-use-for-my-usb-drive/>.
- gmckeown. “Structuring the Records Continuum - Part One: Postcustodial Principles and Properties.” Accessed February 21, 2017. <http://www.infotech.monash.edu.au/research/groups/rcrg/publications/recordscontinuum-fupp1.html>.
- “GNOME System Tools.” Accessed December 6, 2017. <https://projects.gnome.org/gst/>.
- “HandBrake Documentation — Video Encoding Speed.” Accessed October 2, 2017. <https://handbrake.fr/docs/en/1.0.0/technical/video-encoding-performance.html>.
- Han, Yan. “Cloud Storage for Digital Preservation: Optimal Uses of Amazon S3 and Glacier.” *Library Hi Tech* 33, no. 2 (2015): 261–71.
- Harrison, Donald. “Computers, Electronic Data, and the Vietnam War.” *Archivaria* 26 (Summer 1988): 18–32.
- Harvey, Ross, and Martha Mahard. “Mapping the Preservation Landscape for the Twenty-Frist Century.” *Preservation, Digital Technology & Culture* 42, no. 1 (2013): 5–16.
- Hedstrom, Margaret. “A Time Bomb for Digital Libraries.” *Computers and the Humanities* 31, no. 3 (1998 1997): 189–202.
- Hoffman, Chris. “What Lossless File Formats Are & Why You Shouldn’t Convert Lossy to Lossless.” Accessed December 16, 2017. <https://www.howtogeek.com/142174/what-lossless-file-formats-are-why-you-shouldnt-convert-lossy-to-lossless/>.
- “Import Metadata | Documentation (Archivemata 1.6) | Archivemata: Open-Source Digital Preservation System.” Accessed December 3, 2017. <https://www.archivemata.org/en/docs/archivemata-1.6/user-manual/transfer/import-metadata/#import-metadata>.
- “Import Metadata | Documentation (Archivemata 1.6) | Archivemata: Open-Source Digital Preservation System.” Accessed December 3, 2017. <https://www.archivemata.org/en/docs/archivemata-1.6/user-manual/transfer/import-metadata/#import-metadata>.
- “ISO 14721:2012 - Space Data and Information Transfer Systems -- Open Archival Information System (OAIS) - Reference Model.” ISO. Accessed February 21, 2017. http://www.iso.org/iso/catalogue_detail.htm?csnumber=57284.
- Kastellec, Mike. “Practical Limits to the Scope of Digital Preservation.” *Information Technology and Libraries* 31, no. 2 (2012): 63–71.
- — —. “Practical Limits to the Scope of Digital Preservation,” n.d.

- Krahn, Konrad. "Looking Under the Hood: Unraveling the Content, Structure, and Context of Functional Requirements for Electronic Recordkeeping Systems." Master of Arts, University of Manitoba/University of Winnipeg, 2012.
- Lee. "Open Archival Information System (OAIS) Reference Model," n.d.
- Lee, Christopher. "A Framework for Contextual Information in Digital Collections." *Journal of Documentation* 67, no. 1 (2011): 95–143.
- — —. "Archival Application of Digital Forensics Methods for Authenticity, Description and Access Provision." In *Proceedings of the International Council on Archives Congress*. Brisbane, Australia, 2012.
- Lee, Christopher A. "Open Archival Information System (OAIS) Reference Model." In *Encyclopedia of Library and Information Sciences*, Third., 4020–30. Taylor & Francis, 2010.
- — —. "Open Archival Information System (OAIS) Reference Model." *Encyclopedia of Library and Information Sciences*, n.d.
- Lee, Christopher A., Alexandra Chassanoff, Kam Woods, Matthew Kirschenbaum, and Porter Olsen. "BitCurator: Tools and Techniques for Digital Forensics in Collecting Institutions." *D-Lib Magazine* 18, no. 5/6 (May 2012). <https://doi.org/10.1045/may2012-lee>.
- Lee, Christopher A., Richard Marciano, Chien-Yi Hou, and Chirag Shah. "Mainstreaming Preservation through Slicing and Dicing of Digital Repositories: Investigating Alternative Service and Resource Options for ContextMiner Using Data Grid Technology," 113–20. San Francisco, California: California Digital Library, 2009.
- Lee, Christopher A., Porter Olsen, Alexandra Chassanoff, Kam Woods, Matthew Kirschenbaum, and Sunitha Misra. "From Code to Community: Building and Sustaining BitCurator through Community Engagement." A Product of the BitCurator Project, September 30, 2014.
- Lee, Christopher, and Helen Tibbo. "Where's the Archivist in Digital Curation? Exploring the Possibilities through a Matrix of Knowledge and Skills." *Archivaria* 72 (2011): 123–68.
- Lee, Christopher, and Kam Woods. "Automated Redaction of Private and Personal Data in Collections," 298–313. Vancouver, Canada: UNESCO, 2012.
- 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*, no. 72 (2011): 239–46.
- "Maintenance | Documentation (Archivematica 1.4) | Archivematica: Open-Source Digital Preservation System." Accessed October 2, 2017. <https://www.archivematica.org/en/docs/archivematica-1.4/admin-manual/maintenance/maintenance/>.
- 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*, no. 43 (1997): 158–73.
- McDonald, John. "Managing Records in the Modern Office: Taming the Wild Frontier." *Archivaria*, no. 39 (1995): 70–79.
- McDonough, Jerome. "'Knee-Deep in the Data': Practical Problems in Applying the OAIS Reference Model to the Preservation of Computer Games," 1625–34, 2012.
- McLellan, Evelyn Peters. "Selecting Formats for Digital Preservation: Lessons Learned during the Archivematica Project." *Information Standards Quarterly* 22, no. 2 (Spring 2010): 30–33.
- McNally, Tyler. "Practical Digital Preservation: In-House Solutions to Digital Preservation for Small Institutions." *bloggERS!* (blog), January 23, 2017. <https://saaers.wordpress.com/2017/01/23/practical-digital-preservation-in-house-solutions-to-digital-preservation-for-small-institutions/>.
- "md5deep and Hashdeep." Accessed October 2, 2017. <http://md5deep.sourceforge.net/>.
- "md5deep Man Page." Accessed October 2, 2017. <http://md5deep.sourceforge.net/md5deep.html>.
- Meister, Sam, and Alexander Chassanoff. "Integrating Digital Forensics Techniques into Curatorial Tasks: A Case Study." *International Journal of Digital Curation* 9, no. 2 (2014): 6–16.
- "METS: An Overview & Tutorial: Metadata Encoding and Transmission Standard (METS) Official Web Site." Accessed October 2, 2017. <http://www.loc.gov/standards/mets/METSOverview.v2.html>.
- Mumma, Courtney. "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,'" n.d.
- Mumma, Courtney, 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 VANCO_Records AS Archives WHERE Value='true.'" *Archivaria* 72 (Fall 2011): 93–122.
- "MYSQL Issues - Google Groups." Accessed October 2, 2017. <https://groups.google.com/forum/#!searchin/archivematica/mysqlexport:date/archivematica/LeoCgXPQDYI/EssKMd1EOgAJ>.

- Naugler, Harold. "The Archival Appraisal of Machine-Readable Records: A RAMP Study with Guidelines." Paris: United Nations Educational, Scientific and Cultural Organization, November 1984.
- "Normalization for Preservation Issues - Google Groups." Accessed October 2, 2017. <https://groups.google.com/forum/#!searchin/archivemata/mysqllsort:date/archivemata/oC18Puc-nbw/I30VIJlxCWAJ>.
- Oehlerts, Beth, and Shu Liu. "Digital Preservation Strategies at Colorado State University Libraries." *Library Management* 34, no. 1/2 (2013): 83–95.
- "Overview - Archivemata." Accessed March 16, 2017. <https://wiki.archivemata.org/Overview>.
- PREMIS Editorial Committee. *PREMIS Data Dictionary for Preservation Metadata*. 3rd ed. Library of Congress, 2015.
- "Preservica Overview - Preservica." Accessed February 21, 2017. <http://preservica.com/preservica-overview/>.
- "Preservica: The Product Walkthrough - Preservica." Accessed March 16, 2017. <http://preservica.com/resource/preservica-the-product-walkthrough/>.
- "Project Wiki Updates - Google Groups." Accessed April 2, 2017. https://groups.google.com/forum/?fromgroups#!topic/archivemata/K_Sc-h-trms.
- Robertson, Anne, and Adrian Cunningham. "Documenting the Business of Government-Archival Issues in the Digital Age." *Australian Academic & Research Libraries* 31, no. 4 (2000): 188–201.
- Rosenthal, David. "Bit Preservation: A Solved Problem?" *The International Journal of Digital Curation* 5, no. 1 (2010): 134–48.
- Sanett, Shelby. "Archival Digital Preservation Programs: Staffing, Costs, and Polcy." *Preservation, Digital Technology & Culture* 42, no. 3 (2013): 137–49.
- Schumacher, Jaime, Lynne Thomas, and Drew VandeCreek. "From Theory to Action: 'Good Enough' Digital Preservation Solutions for Under-Resource Cultural Heritage Institutions." Illinois: Institute of Museum and Library Services and POWRR, 2014.
- Scott, Bethany. "Preservation and Access Can Coexist: Implementing Archivemata with Collaborative Working Groups." *blogGERS!* (blog), February 28, 2017. <https://saaers.wordpress.com/2017/02/28/preservation-and-access-can-coexist-implementing-archivemata-with-collaborative-working-groups/>.
- "SIP Structure - Archivemata." Accessed October 2, 2017. https://wiki.archivemata.org/SIP_Structure.
- Sprout, Bronwen, and Mark Jordan. "Archivemata As a Service: COPPUL's Shared Digital Preservation Platform." *Canadian Journal of Information and Library Science* 39, no. 2 (June 2, 2015): 235–44.
- Taylor, Hugh. "Information Retrieval and the Training of the Archivist." *The Canadian Archivist* 2, no. 3 (1972): 30–35.
- Tibbo, Helen, and Christopher Lee. "Closing the Digital Curation Gap: A Grounded Framework for Providing Guidance and Education in Digital Curation." Society for Imaging Science and Technology, 2012.
- Tilbury, Jon. "Avoiding a Digital Dark Age: The Governance of Long-Term Digital Information." *Information Today*, April 2016.
- Trace, Ciaran. "Beyond the Magic to the Mechanism: Computers, Materiality, and What It Means for Records to Be 'Born Digital.'" *Archivaria* 72 (Fall 2011): 5–27.
- "TRAC Metrics | CRL." Accessed February 21, 2017. <https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac>.
- "TRAC Metrics | CRL." Accessed October 3, 2017. <https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac>.
- "University of Manitoba - Libraries - Mission and Mandate." Accessed October 3, 2017. <http://www.umanitoba.ca/libraries/units/health/archives/quickfacts.html>.
- Upward, Frank, Sue McKemmish, and Barbara Reed. "Counterpoint: Archivists and Changing Social and Information Spaces: A Continuum Approach to Recordkeeping and Archiving in Online Cultures." *Archivaria* 72 (Fall 2011): 197–237.
- Van Garderen, Peter. "Archivemata: Using Micro-Services and Open-Source Software to Deliver a Comprehensive Digital Curation Solution." In *Proceedings of the 7th International Conference on Preservation of Digital Objects*, 145–49. Austrian Computer Society, 2010.
- "Vault7 - Home." Accessed March 10, 2017. <https://wikileaks.org/ciav7p1/>.
- "Virtual Appliance Instructions - Archivemata." Accessed October 2, 2017. https://wiki.archivemata.org/Virtual_appliance_instructions.
- "Wayback Machine." Accessed April 2, 2017. https://web-beta.archive.org/web/20120901000000*/http://archivemata.org.

- “Wayback Machine.” Accessed April 2, 2017. https://web-beta.archive.org/web/20100601000000*/http://artefactual.com.
- “What Is Web 2.0.” Accessed April 2, 2017. <http://oreilly.com{file}>.
- Wiedeman, Gregory. “Python for Archivists: Breaking down Barriers between Systems.” *Practical Technology for Archives* (blog), December 20, 2016. https://practicaltechnologyforarchives.org/issue7_wiedeman/.
- Wilson, Andrew. “How Much Is Enough: Metadata for Preserving Digital Data.” *Journal of Library Metadata* 10 (2010): 205–17.
- Yeo, Geoffrey. “‘Nothing Is the Same as Something Else’: Significant Properties and Notions of Identity and Originality.” *Archival Science* 10 (2010): 85–116.
- Zaste, Chris. “Another Bit Bytes the Dust: The Technological and Human Challenges of Digital Preservation.” Master of Arts, University of Manitoba/University of Winnipeg, 2016.