Above and Beyond the File:

Preserving Digital Environments

By

Andrew Webster

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 Master's Program

University of Manitoba/University of Winnipeg

Winnipeg, Manitoba

Table of Contents

Abstract

Digital Record formats have various dependencies tied to hardware and software that must be fulfilled if a record is to be rendered and therefore accessible. This relationship has proven difficult to preserve due to the fragility and obsolescence of digital records, media, and environments. In response to this challenge a consensus has formed within the digital preservation community around format management through format migration. To remain ahead of obsolescence and move away these dependencies digital records are often migrated to more open formats that are more platform independent. Format migration often does not preserve the original functionality of formats, which can also present difficulties rendering data, and removes the migrated record from the original context in which it was created. This raises two questions, which will be answered in this thesis: first, what are the effects of failing to preserve digital environments; and second, how might we go about preserving them?

This thesis will examine, in addition to format managements, several alternative methods of preservation aimed at preserving digital environments and communicating original record-environment relationships to users. These include: preservation of legacy environments; emulation, including efforts to modern combine cloud computing power with the "as-a-service" model to enable widespread adoption of emulation; and, expanding upon existing archival descriptions by including existing data and materials and supplementary documentation. This thesis will conclude with a case study of the William O. Pruitt Jr. fonds at the University of Manitoba Archives and Special Collections (UMASC). This fonds was selected as a case study as it contains digital records from the early era of desktop computing. This offers a good example of what is gained and lost when preserving these records through format management, which is the approach to digital preservation followed by UMASC, or if preserved through

environment-based methods such as emulation. Interviews were conducted via questionnaire with the University of Manitoba's Libraries Research Services & Digital Strategies team (RSDS), so that they could explain their chosen mode of digital preservation and the specific strategies that they follow, including their use of Archivematica.

Acknowledgments

I would first like to thank Professor Tom Nesmith and the energy and passion he put into his Introduction to Archives course I attended during my undergrad. It was through this course that I truly gained an appreciation for the world of archiving and further cemented while attending his lectures during the Joint Master's Program.

Of course I must also profusely thank Professor Greg Bak. First, for the boundless energy (and last-minute supplementary readings) he instilled into teaching during the Joint Master's Program. Second, for his guidance throughout the writing of this thesis helping to ensure I could produce coherent thoughts on paper.

I must of course acknowledge the two Participants who responded to the questionnaire. I thank them for their time and professional input.

I would also like to thank my family for the support over the years. First, when hearing I was switching pursuits from engineering to history, and second, when I later proclaimed the intent to become an archivist. Thank you for letting me pursue what I wished and supporting me through the happy periods and the stressful ones.

Much thanks to my fellow classmates for: their aid during group work; helping bounce ideas off one another during and after class; and the times when it became necessary to refuel on pizza and pop.

Introduction

Digital technologies have become an essential part of our working and personal lives. They also present unique challenges with regards to preservation. While records and materials in all forms have unique requirements associated with their preservation digital records stand out due to their technological dependencies. Obsolescence is a greater threat to digital records and materials than traditional paper since they are more complex and cannot be viewed on their own but must be rendered in a current digital system and displayed on a computer interface. Such interfaces are digital environments, which for the purposes of this thesis are defined as the hardware and software required to properly interpret the bit sequences of a digital record and render it for human viewing and use allowing for interaction and potentially modification.[1] This intermediary step of rendering records is in fact made up of several processes all dictated by dependencies upon specific hardware and software. Unfortunately, it is difficult to preserve digital environments owing not only to their complexity but also their fragility which is particularly evident in the increasing scarcity of hardware from the 1980s and early 1990s. Many records from this era are the most at-risk since they often reside on obsolete storage media such as floppy disks. Many file formats and the software necessary to render them are also no longer in use or have been replaced by more up-to-date versions that may have little-to-no support for backwards compatibility. The consensus in digital preservation is therefore to focus upon format management in order to work around the issues of software and hardware dependencies.

Preserving files by migrating them to different formats so that they are less dependent upon specific environments, and therefore more widely accessible, often comes at the cost of functionality. The issues of compatibility also mean that some digital records may be preserved

[1] Angela Dappert, Sébastien Peyrard, Carol H. Chou, and Janet Delve, "Describing and Preserving Digital Object Environments," *New Review Of Information Networking* 18 no.2 (November 2013): 110-11.

but remain relatively inaccessible. There is also the fact that the environment is also often absent in the transfer of records to archives. It is not that the environment has been deemed to lack archival value, rather, it is simply much easier to make the contents of records accessible though format migration. Donors are also either unwilling or unable to part with original environments since they may still be using them or no longer have them. Knowledge of and access to obsolete environments (also referred to as "legacy" environments throughout this thesis) is essential to digital preservation work since they may hold records or be required to access legacy media. This does not mean it is impossible to preserve digital environments, or at the very least aspects of their functionality. It is also possible to develop supplementary materials to include as reference with existing digital records. More importantly, computers have grown powerful and sophisticated enough that tools that were once only hypothetical solutions to preserving environments have potentially become more practical. With this in mind there are two major questions this thesis will endeavour to answer: first, what is the effect of failing to preserve digital environments; and second, how might we go about preserving them? In order to answer the first question it will be important to explore the historical and technical significance of digital records and environments. As Michael S. Mahoney notes, the history of the computer has primarily been about the machine, its sudden appearance, and seemingly inevitable evolution into the modern desktop workstation. He argues this is an overly simplistic model and ignores much of the contextual history, specifically the agency of people and how they impacted and shaped technology, rather than acting in a purely reactionary manner. "The devices and systems of technology are not natural phenomena but the products of human design, that is, they are the result of matching available means to desired ends at acceptable cost".[2]

[2] Michael Mahoney, "Histories of Computing(s)," *Interdisciplinary Science Reviews*, 30 no. 2 (2005): 121-2.

Chapter One will address the influences that computers and users have had upon each other over the last few decades to provide historical and technical context. First will be an examination of the nature of digital records and the significance of their inherent manipulability and increased functionality with regards to change and authenticity compared to traditional paper records. This will include exploring the significance of recordkeeping systems and how they guide the creation of records and define their purposes. Such systems, both paper and digital, act as higher level environments that guide and define the tasks of whole institutions, departments, and staff. Digital environments associated with specific types of computing infrastructure, such as mainframes through to desktop computers offer more granular examples of environments. The focus of this chapter will then shift to cover the introduction of these systems into the working and public spheres. This will include the expectations and reservations regarding automation in records production and management and concerns about their archival value and the role of archivist in their preservation.

The expectations and realities of what automated computing could accomplish and debate regarding the value of digital materials and role of archivists evolved as computers became more prevalent. It is here that the matter of agency will be most apparent. Early perceptions of computers in the 1940s and 1950s centered around the potential for what some called a second industrial revolution and mainframes became seen as symbols of modernity and pride for the few institutions that could afford them. Moving into the 1960s and 1970s when computer adoption was more widespread, but still limited to businesses and government institutions, debate was centered around whether it was practical to preserve digital records. If so, was it then the responsibility of archivists to also develop the technical skills necessary to understand and operate computer systems or simply preserve the records and leave such operation to users. The

1980s and in particular the 1990s saw discussion focus on how best to preserve digital data and whether environments were necessary beyond the transfer process or if alternatives such as emulation should become the future standard of digital preservation. The expansion of computer adoption also spurred movements within society from protestors concerned that people were being treated as statistics rather than individuals to the evolution of early electronic hobbyist movements that originated with radios. The purpose of this chapter is to examine the technical and cultural significance of environments by exploring the cultural significance of computing and meditating upon the location of archival value within digital systems and records.

Chapter Two will focus upon four methods of digital preservation including format management and migration, around which there is a consensus among most working digital archivists. Alternatives to format management that will be considered include: the physical preservation of legacy environments; utilization of emulation which allows select traits of older environments to be accessible in modern ones; and combining description and supplementary documentation to represent the missing functionality of absent environments for users. The aim of this chapter is to address the second question of how we might go about preserving digital environments. This chapter will open by describing why migration-based format management has become the consensus within the field of digital preservation and the current role of legacy environments therein. It will then expand upon this latter point to explore the benefits and issues concerned with preserving legacy hardware. Emulation, the reproduction of an environment solely as software, has been touted as the ultimate solution to obsolescence. It is the most technical and resource intensive method of preservation encompassing multiple disciplines including archiving and computer science. Currently efforts are being directed towards combining emulation and cloud computing to develop Emulation-as-a-Service (EaaS) as a means

to centralise resources and expertise to offset the costs to potential users. The fourth and potentially most practical method is not to directly preserve environments. Instead it focuses upon leveraging existing technical information and alternative sources to collect and create supplementary documentation addressing the significance of the environment in the creation of the original records and their functionality. The aim is to expand existing archival description and other methods of description and documentation to better accommodate environment related information so that it can be made available to users.

Chapter Three will focus upon a case study involving legacy digital records and media from the University of Manitoba Archives and Special Collections (UMASC) holdings, the William O. Pruitt Jr. fonds. This chapter will also incorporate responses by members of the University of Manitoba Libraries Research Services & Digital Strategies (RSDS) team, who are responsible for digital preservation, to a questionnaire regarding their work.[3] The Pruitt fonds is ideal since it was a clearly defined set of records including detailed descriptions of the contents of the disks and has already been submitted to the appraisal process and deemed to be of archival value. The purpose of this exercise is not to question the archival value of these specific records, but to explore how it may be possible to better leverage the existing value in the records through digital preservation methods beyond format migration alone. UMASC and RSDS follow the current consensus around format management as a means for making digital records available to users, which also made them ideal candidates for the case study. While RSDS does maintain a blog concerning the work they do, which provides some insights into their processes and decision-making, the interviews were an opportunity to go beyond the published works and obtain a greater understanding of their work.

[3] The Questionnaire and anonymized responses from both Participants can be viewed in full in the Appendix at the end of this thesis. Both Participants chose the option to remain anonymous in accordance with the University of Manitoba Research Ethics Board.

By not preserving digital environments or better communicating original functionality to users, they will be unable to comprehend the full scope and experience of digital records and their technical provenance which includes their functionality. They also cannot comprehend or appreciate the efforts archivists must undertake to preserve the records in order to make them accessible. Understanding and appreciating these functionalities is essential for users to begin to comprehend the capabilities and limitations experienced by the creators of digital records. Like Mahoney, Mathew Connell has examined the historical significance of computer systems, noting that "It was not so much a matter of 'how it works' but 'what it does'."[4] Preserving digital environments is not limited solely to preserving technical specifications, it is also a means to preserve the agency of the creators.

[4] Connell, "Computer as Historical Artefacts," *Australian Journal of Multi-Disciplinary Engineering*, 4 no. 1 (2006): 41.

Chapter One – The Digital Environment and History of Digital Preservation

Introduction to Digital Environments

Digital records pose several unique problems when it comes to preservation, particularly in relation to format and function. There are myriad of different formats and related requirements to view, edit, and interpret digital records, which can be difficult if the necessary hardware and software are worn out or scarce due to obsolescence. This has led to the development and adoption of various tools and methods dedicated to managing digital records and ensuring long-term accessibility. One of the most popular methods is format migration which can be summarised as "the process of creating a version of a file in a more current format, particularly if the format of the source file is in danger of becoming obsolete. Ideally, format migration should be as lossless as possible and retain the content, appearance, and behaviors of the source."[1] There are two important points to take away from this definition: first, it results in the creation of a new version of a file; and second, it ideally relies upon lossless format migration. This means that it is possible to change the record *itself* without the need to rewrite (or type) the contents from scratch in a new document, but such an act may run the risk potential loss of form and function during this process. Arguably, such loss would mean that the migrated version of the record is to some extent damaged.

If there is the potential for any aspect of a record to be poorly adapted, or even lost, then why undertake such a process? As Yvette Hackett bluntly states: "digital technology is a moving target … upgrading constantly while simultaneously becoming obsolete at the same rapid pace."[2] Obsolescence has much greater consequences for digital records than their traditional non-digital

[1] Priscilla Caplan, "The Florida Digital Archive and DAITSS: A Working Preservation Repository Based on Format Migration." *International Journal on Digital Libraries* 6 no. 4 (2007): 307.
[2] Yvette Hackett, "Preserving Digital History: Costs and Consequences," in *Better off Forgetting?: Essays on Archives, Public Policy, and Collective Memory* ed. Mona Homlund and Cheryl Avery (Toronto: University of Toronto press, 2010): 124.

counterparts. Though it may have faded the cover art and loosened the binding of a book from the 1980s it can still be read in its entirety by anyone simply by picking it up and opening it. For digital records of a similar age it can be much more frustrating. The contents on a 5.25-inch floppy disk cannot be read without the appropriate drive, and if it is even possible to get hold of one it is very likely not supported by most modern computers. Adrian Brown, in his book *Practical Digital Preservation: A How-To Guide for Organizations of Any Size*, states that format migration presents preservationists with the option to "transform the original object to a form that is no longer reliant on obsolete technology, but can instead be accessed using whatever is standard at the time."[3] With the rapid pace at which digital technology develops, keeping digital records "current" and accessible over time are major points of concern for any institution with a focus towards the long-term preservation of digital records. This may even necessitate the creation of newer migrated versions in more up-to-date formats, even replacing older ones due to practical concerns such as storage space. Brown firmly states that while such replacement may be necessary "the relationships between manifestations, as recorded in the technical metadata, must never be destroyed: the authenticity of a record requires the maintenance of an unbroken chain of provenance." It is also important to note that format migration is not meant to replace original versions; those must be preserved to ensure an unaltered version of a record exists, to ensure the integrity of the record.[4] This is arguably the most important link in the chain of provenance between manifestations. The potential to migrate away from dependency upon older digital environments is one of the greatest benefits of working with digital records and aids archivists in their efforts to keep pace with the digital "moving target", as Hackett would put it.

[3] Adrian Brown, *Practical Digital Preservation: A How-To Guide for Organizations of any Size* (London: Facet Publishing, 2014): 209.
[4] Brown, *Practical Digital Preservation*, 217-18.

This ability to reformat[5] the record itself, or at least make a more accessible version, is impressive. However, while changing the record may not necessarily change the content itself, it does change the context in which it is presented. Migrating a record to a different format means removing it from the environment it was created in and presenting its contents in a different and more modern one. As this thesis focuses upon the importance of the relationship between digital files and environments it is important to have a definition of an environment. Digital environments are the hardware and software required to properly interpret the bit sequences of a digital record and render it for human viewing and use allowing for interaction and potentially modification.[6]

The Consensus – Format Migration

As mentioned in the introduction, a digital record must first be rendered in a way that is understandable to humans in order for them to be able to interact with it. By comparison, a book requires no such additional interpretation in order to be made accessible. Even if a user does not understand the language within a book they can still turn the pages and view the contents. If the bit sequences of a digital record cannot be properly interpreted, the record cannot be rendered and, therefore, it does not matter if the user cannot understand the language the contents are written in since they are unable to even view it. Format migration offers a relatively straight-forward solution by transforming records into formats that are more "open" and stable, meaning they have fewer dependencies required to render them. This allows for the creation of versions of formerly obsolete records created in legacy environments which can be read and rendered using "Whatever is standard at the time."[7]

[5] To be accurate, reformatting means creating a whole new record altogether.
[6] Angela Dappert, Sébastien Peyrard, Carol H. Chou, and Janet Delve. " Describing and Preserving Digital Object Environments." *New Review Of Information Networking* 18 no.2 (November 2013): 110-11.
[7] Brown, *Practical Digital Preservation*, 209.

As with non-digital records and materials one of the most difficult parts of digital preservation is handling the volume of materials, not to mention the variety of storage mediums and formats, all of which are subject to the threat of obsolescence. As an example, Library and Archives Canada (LAC) currently handles over thirteen petabytes of digital content in their Digital Archives and Preservation Centre which, without long-term preservation measures that include access, "will become completely inaccessible and are at risk of becoming permanently lost."[8] These records were created over the last half century on several generations of computing systems. Given the quantity and time period format migration is likely the only efficient method of preservation that will allow access. Original equipment is either gone or too varied to preserve since this would include several generations of distinct computing environments. Efforts to develop emulated environments as alternatives would simply be unfeasible. The rapid increase in the volume of digital records entering archives over the past thirty years also exacerbates the issue. A 2014 report by the Auditor General noted that despite stating digital records would be the "format of choice" by 2017, at the time of the report LAC did not have a sufficient corporate digital strategy in place to address long-term digital preservation goals. This included establishing clear terms and conditions regarding "necessary metadata or institution-specific format requirements for the transfer of digital records."[9] In response, LAC began an ongoing overhaul in terms of policy and practice, including the development of a new digital asset management system aimed at automating and better coordinating ingesting, cataloguing, and preserving digital content. This included working closer with government departments to

[8] "Strategy for Digital Preservation," Library and Archives Canada, Date Last Modified, December 31, 2020, https://www.bac-lac.gc.ca/eng/transparency/briefing/2019-transition-material/Pages/digital-strategy-digital-preservation.aspx.

[9] Canada, Auditor General, "Chapter 7—Documentary Heritage of the Government of Canada—Library and Archives Canada," in *2014 Fall Report of the Auditor General of Canada*, 2014, https://www.oag-bvg.gc.ca/internet/English/parl_oag_201411_07_e_39965.html#appa.

establish criteria for preparing records for transfer to LAC including aid in aligning their systems to accommodate this and better communicate with LAC's system. [10]

While the quantity of records at LAC makes it atypical among Canadian archives, the fundamental challenge remains the same. As at LAC, most archives have significant quantities of digital records already in their vaults, accessioned along with non-digital records over the past four or five decades. This includes the University of Manitoba Archives & Special Collections (UMASC), which is of particular interest for this thesis on account of the case study in Chapter Three. As at LAC, the archivists and digital preservation specialists at the University of Manitoba identified format migration as the best means of preserving and making accessible these digital records. Focusing on format migration allows archivists to deliver to users digital records, now, that can be rendered and accessed on current computer systems. These records can be made available in the archives' reading room or by sharing them over a network such as the Internet. This flexibility and the promise of immediate accessibility has allowed a consensus to form among Canadian archivists that format migration is the best means of addressing the challenges posed by digital records. In selecting format management as their primary method of preserving digital records and making them available LAC and UMASC are completely typical examples of this consensus.

Format migration takes advantage of the technical fluidity of most digital records and benefits from the fact that it can be, to varying extents, automated. Utilizing migrated versions which are readily accessible on modern computer systems and can be produced in large volumes are why format management, utilizing migration, has become the consensus within the field of digital preservation. Format migration is also not limited to combating issues of obsolete

10 Sylvain Bélanger, "LAC Digital Optimization of Canada's Collections," Presentation, University of Manitoba, December 2, 2019.

formats. Many archives perform migrations to multiple modern formats in effort to provide users with choices regarding how they consume digital content.[11] Format management enables archivists to meet the needs of modern users and computer systems with regards to accessibility of digital content originally residing solely within obsolete formats.

Nonetheless, while format migration may be the consensus it has not gone without criticism. In the 1990s Jeff Rothenberg argued that format migration (or "translation" as he put it) was a flawed method of long-term digital preservation. He argued that, while translating digital records to new formats may seem similar to translating ancient literary works into modern languages, digital records had notable limitations. Translation between formats also introduces losses such as the displacing of content and alterations in how it is structured which are compounded by the fact that new formats develop rapidly and often lack compatibility with older ones. Unlike literary works, such loss is often irreversible because even though the original record is preserved obsolescence means that that original software may not, or cannot, be run.[12] A straightforward example would be a Microsoft Office Word document. While the current version of Word can access some documents in older formats (.doc rather than the more recent .docx) they will likely be rendered differently in the newer version since symbols, fonts, and other features and functions can change or be removed. This can lead to blank spaces, the rearrangement of tables and images, and random symbols appearing in place of letters. Metadata may not be present within the migrated format and embedded hyperlinks will break or not register. While this may seem to be merely a cosmetic matter the implications can potentially run much deeper, perhaps obscuring the thought processes of the creator. Chapter Two will discuss

[11] Brown, *Practical Digital Preservation*, 206.
[12] Jeff Rothenberg, "Ensuring the Longevity of Digital Documents," *Scientific American* 272 no. 1 (1995): 44-6.

how some of these issues are circumvented such as through digital forensics and the ability to extract metadata from legacy files to help identify them and possible migration paths.

Rothenberg argued that the focus should be on preserving software and the necessary contextual information that would allow future users and more powerful computers to accurately replicate its behaviour through emulation.[13] Much of Rothenberg's vision has in fact, to some extent, come to pass. Computers have grown more powerful allowing for the emulation of whole computing systems and generating and collecting technical metadata detailing the migration process is standard practice. These are all important to developing and preserving the chain of provenance, as Brown emphasises. To extend Brown's metaphor, the environment is a part of this chain since through it the record is first created and manipulated before being translated. Without it the chain is incomplete. Provenance is more complex than simply listing pathways so while Brown's definition is useful it is also limiting in this regard. What is required is a more encompassing definition of provenance is required going forward.

An alternative definition of provenance can be found in the writing of Tom Nesmith: "The provenance of a given record or body of records consists of the social and technical processes of the records' inscription, transmission, contextualization, and interpretation which account for its existence, characteristics, and continuing history."[14] Christoph Becker, Director of the Digital Curation Institute, and Associate Professor at the University of Toronto, expressed similar concerns, though from a computing and systems design perspective. In his words "damage occurs often not as a loss of physical integrity, but as a loss of relationships between

[13] Rothenberg, "Ensuring the Longevity of Digital Documents," 47.
[14] Tom Nesmith, "Still Fuzzy, But More Accurate: Some Thoughts on the "Ghosts" of Archival Theory," *Archivaria* 47 (1999): 146.

elements, whether through link rot, obsolescence, or lack of metadata."[15] Including relationships alongside their respective records as part of the preservation process allows users to better grasp the significance of the record and how it was created. Failure to include the file-environment relationship, and instead relying solely upon changing formats, can be detrimental to digital preservation. The absence of the environment itself or reference to it means that, while users will be aware the records they are viewing were created on older computer system, it may not be readily apparent to user what specific system or software was used. Becker emphasises that digital preservation is about more than keeping the digital record safe from harm. It is important to acknowledge that "Electronic records were always computational products of algorithms that used data as input to create a performance."[16] A digital record is more than a static object, it is a combination of processes that serve to interpret and display data in some meaningful way. As computing historian Michael S. Mahoney stated, computers "take sequences, or strings, of symbols and transform them into other strings … in the end, computation is about rewriting strings of symbols … [and] any meaning the symbols may have is acquired and expressed at the interface between a computation and the world in which it is embedded."[17] This performance, the combination of processes and transformations, is more than a means to access a record, it is a part of the record itself.

What Change Means for Digital Records

As stated by Becker and Mahoney, computing is all about transforming and rewriting information in order to present it in a manner which users can view and interpret. Then if computing is all about transformation, why is format migration a problematic method of digital

[15] Christoph Becker, "Metaphors We Work By: Reframing Digital Objects, Significant Properties, and the Design of Digital Preservation Systems," *Archivaria* 85 (2018): 29.
[16] Becker, "Metaphors We Work By," 31.
[17] Michael S. Mahoney, "The Histories of Computing(s)," *Interdisciplinary Science Reviews* 30 no. 2 (2005): 129.

preservation? The concern lies in the fact that format migration is primarily content driven with regards to preservation. Migrating a record to a different format often means that much or all of the original functionality is lost through such transformation. For future users who do not share the familiarity with the materials or transformation process that archivists posses it may not be readily apparent that the record has been transformed or to what extent it has altered the record.

Format migration merely transfers some (but not all) of the contents of a record from one format to another so it may be accessible on other environments. Unless *explicitly* documented, through human intervention or automated metadata generation, any evidence of the technical and social processes involved are left behind. Becker points out that ultimately, the outcome of digital preservation should be about reproducing an authentic performance from the record, such as how it is rendered on a screen.[18] A straightforward example is using format migration to preserve the contents of an Excel spreadsheet to a PDF for ease of access. Users can still view the outline of the cells and the visible contents. However, the formulas that are part of the core functions within Excel that allows for the manipulation and generation of data are gone. Rather than a full performance this is more akin to a snapshot. The result is incomplete information and a disjoint between the migrated content and its original context, all of which is damaging to the record. This is the "loss of relationships between elements" Becker alluded to. This is why the method of transformation is such an issue for digital records because both the content and the way in which it is presented can be negatively affected by format migration. The formulas may not be visible at all times in the original spreadsheet format but they are accessible to users and informed users can use them to determine the (multiple) functions they served. Changing the format is a major event in the history of a record because this recontextualizes how it is presented going forward within more modern environments rather than the environment it was

[18] Becker, "Metaphors We Work By," 23.

created in. This means that the record will be interpreted differently going forward. This makes format migration an example of the technical and social processes Nesmith mentions (though not one he specifically discusses). While it may be a necessary process it is important to understand that changing the record to preserve access to it can also be damaging.

When a record or artifact experiences some form of physical alteration this can lead many to question its validity going forward, even so far as to argue it can no longer be considered the original. Heather MacNeil cites the controversy surrounded the cleaning of the ceilings of the Sistine Chapel between 1984 and 1994 as a non-digital example of how physical change to an artifact can cause division within the preservation community; supporters claimed they were removing the "grime of the centuries" and returning it to its original form just as Michelangelo expected it to be viewed when he painted it; critics argued that such "grime" was an essential part of the frescoes and the source of their historical authenticity as a work of a master.[19] The removal of "grime of the centuries" was an effort to improve accessibility to both the frescos themselves and what supporters believed to be the intentions of the creator. To critics this was an act of removing what had essentially become a physical and historical part of the frescos themselves. The debate concerned form and function. Did the grime interfere with the ability of viewers to see the work as the artists had originally intended, or was it a significant indicator of authenticity literally being erased through the cleaning process?

An example closer to home would be the staining of the *Proclamation of the Constitution Act, 1982* in July of 1983. Peter Greyson entered the Public Archives and National Library Building on the 22nd, using his student identification from the Ontario College of Art, expressing a desire to view the *Proclamation* for research purposes. He was then witnessed pouring red paint from a container he had hidden in his pocket in what he proclaimed was an act of protest

[19] Heather MacNeil, "Archivalterity: Rethinking Original Order," *Archivaria* 66 (Fall 2009): 6-7.

against then Prime Minister Trudeau's decision to allow the United States to test cruise missiles in Canadian air space.[20] In a matter of seconds the copy of the federal document had undergone a significant transformation. Restoration became a matter of debate since the lead-based pigment could not be removed by most techniques. There was even a suggestion of cutting the stained section out and replacing it, until it was decided that "It [the stain] is part of the history of the thing and shouldn't be taken away."[21] The stain also had farther reaching consequences than the immediate document. At the time public and media attention were drawn towards archival matters such as conservation. The staining has become more than a moment in the *Proclamation's* history, it is now a defining trait giving it a degree of infamy among historical government documents. MacNeil identifies restoration as a form of discontinuous transience; change that is the result of human involvement which "irretrievably alters and reconfigures the work."[22] In the case of the *Proclamation* cutting out and replacing the stained section would have left an obvious physical deformation. Such an alteration would arguably equal to the stain in magnitude.

Format migration is also a form of discontinuous transience. Such change is done in an effort to improve accessibility. However, there are two major differences when it comes to discontinuous transience and digital records. First, is that format migration is usually pre-emptive in order to stay ahead of obsolescence, while the restoration of the Sistine Chapel frescos occurred after they had already deteriorated. This means that digital format migration also serves as a proactive means of preservation whereby an altered version of a record is created before its accessibility is threatened, not afterward. The second difference is the most important due to the

[20] James M. Whalen, ""Out, Damned Spot!" The Staining of the Proclamation of the Constitution Act, 1982," *Archivaria* 61 (2006): 291-3.
[21] Whalen, "Out, Damned Spot!," 295.
[22] MacNeil, "Archivalterity," 7.

issue of functionality. Critics and supporters have debated over what may or may not have been lost with the restoration of the frescoes, whether they have been returned to their intended glory or are no longer truly authentic. This is primarily an aesthetic rather than functional debate since the frescos are still frescos and remain within the church in which they were painted, depicting the same content in the same medium, which leaves much the interpretation up to personal aesthetic. Grime may obscure the content of a fresco, but it does not inhibit its ability to function as a fresco.[23] When a record is migrated to a different format it is bound by different functional limitations. A Word document and a PDF are noticeably different in how they appear when rendered and to what extent the content can be manipulated.

When it comes to migrated versions of digital records, it is often easy to identify exactly what has been lost because the functionality of the record is unequivocally altered or reduced. A PowerPoint presentation may incorporate visual and audio effects including embedded videos or even hyperlinks. Printing the slides or converting them into a PDF to allow users to quickly review them outside of the digital environment – in this case the PowerPoint software – removes any and all interactive function from the presentation. Any sound and video have obviously been abandoned and some content may no longer be visible since it is tied to manual or automatic triggers. It would be like purchasing a photograph of the Sistine Chapel frescos to view at home removed from its environment. The photograph does not function like the original frescos as it is a limited replication. The impression given by the scale of the church and tone set by surrounding religious art, architecture, and iconography are absent. However, as Brown noted, format migration is used to provide users with multiple accessible formats to meet their needs for the consumption of content. Quality format migration is a technically-involved undertaking since it may often require identifying migration pathways that require going through multiple formats.

[23] It is arguable whether or not it limits the impact it has as a centerpiece in the Vatican though.

Testing of pathways is essential and may require visual comparisons of original and migrated versions to ensure unforeseen or undesired alterations have occurred.[24] The example highlights just how significant an impact format migration can have in some circumstances.

Matthew Connell identifies software as contextual material since it is through software that we can make computers perform tasks,[25] tasks which include the creation and manipulation of digital records. "Ones and zeroes can be words, pictures, motion picture or hyper media (combination of media), or program code. Often the only way to retrieve the full meaning of that data is to play it in the context of the document that confirmed it."[26] Format migration focuses on changing the record to accommodate current computing standards in effort to present content in a consistent manner within the context of modern environments. It offers a workaround to reliance upon aging environments and a way to bring raw data and content forward, but at the expense of relevant context and functionality. However, the migrated version does not readily communicate the absence of any functionality granted by the original format and environment. Mahoney makes the profound argument that it is important to understand design as it relates to computers and programing. He specifically points to understanding how a portion of the world is modeled in the computer and its programming so that it can "understand" this portion to a certain extent.[27] This is of course determined by human agency via the needs of users and developers.

By understanding the computer, we can understand social and technical processes emphasized by Nesmith. Ciaran B. Trace notes how the user-computer relationship has changed over the decades from the hobbyist desire to open the computer up and analyse each component,

[24] Brown, Practical Digital Preservation, 210.
[25] Matthew Connell, "Computers as Historical Artefacts," *Australian Journal of Multi-Disciplinary Engineering* 4 no. 1 (2006): 41.
[26] Connell, "Computers as Historical Artefacts," 42.
[27] Mahoney, "The Histories of Computing(s)," 128-9.

to the modern one where interaction is limited to the interface on the screen.[28] A lot of modern technological development and promotion is driven by the promise of content being "accessible anywhere, at anytime", which also translates into minimal user input outside of the surface experience. This also dictates how businesses and institutions, not only archives, make their services available to users, content should be easily accessible with just a few clicks and work on whatever device they are using. Format migration is useful in this regard by moving data to more open formats, but simplifying the user experience cannot be done without understanding certain fundamentals. Trace argues that providing a simplified experience should not be the sole driver of digital preservation. By taking a broader approach archivists can "more fully embrace sophisticated models of how digital records exist within computer systems … [and] … will be better equipped to understand the tools and processes of digital forensics that can help in the acquisition and preservation of born-digital records."[29] The needs of records and users are very different and subject to change over time, a fact which is well known within the field digital preservation given the rapid development of computer systems, software, and formats. Therefore, by attaining even greater understanding of the records and artifacts in their care archivists will be better equipped to adapt to these changes.

Like any other technological achievement, computers were not created in a vacuum but to fulfill a purpose. Mahoney states that "Its nature is protean; the computer is – or certainly was at the beginning – what we make of it (or now have made of it) through the tasks we set for it and the programs we write for it."[30] If the point of preserving records and artifacts is to study and critique the past then it is only logical that the environment should in some way be preserved as

[28] Ciaran B. Trace, "Beyond the Magic to the Mechanism: Computers, Materiality, and What It Means for Records to Be "Born Digital"," *Archivaria* 72 (2011): 6.
[29] Trace, "Beyond the Magic to the Mechanism," 8.
[30] Mahoney, "The Histories of Computing(s)," 122.

well, either directly or through emulation or even dedicated documentation. Looking back to try and understand cultural dynamics of institutions and businesses will require understanding then-modern tools so we may better grasp their capabilities and limitations. This can be difficult to communicate to users when performing more limiting format migrations such as the Excel-to-PDF or PowerPoint-to-printout methods of preserving records. Users see some of the data, but not the performance. Nor can users properly determine agency regarding the capabilities and limitations afforded to the record creators. The volume of records and complexity of how they are ordered can prove to be an obstacle to performing more complex format migrations that can better communicate this to users. An alternative such as emulation, discussed in Chapter Two, has the potential to afford users of archives the opportunity to interact with records in their original formats in addition to migrated versions.

Before the Machine-Readable Record - The Paper Environment

Looking beyond the records to the environments in which they were created and utilized can yield greater insight into what function they served by understanding the relationship between the two. David Bearman states that recordkeeping systems are defined by "the role they play in providing organizations with evidence of business transactions (by which is meant actions taken in the course of conducting their business, rather than 'commercial' transactions)."[31] In this regard a recordkeeping system is more than a set of rules telling users what folder to store records in. It is in fact dynamic, serving as a series of checks and balances that validate and reflect the functions of an organisation. He further elaborates by saying:

> Record-keeping is a critical function that is performed through the collective action of individuals and systems throughout all organizations. Record-keeping is not the

[31] David Bearman, "Record-Keeping Systems," *Archivaria* 36 (1993): 17.

> province of archivists, records managers, or systems administrators alone, but an essential role of all employees and of individuals in their private lives.[32]

Recordkeeping is a collaborative effort that is built upon connections between people and records. Therefore, a recordkeeping system is potentially the greatest set of tools archivists and users of archives can use to understand the creation and intended function of the record. Tom Nesmith provides a succinct description of what an archive is, which is helpful in determining how they may fit into Bearman's model of recordkeeping systems: "An archives is an ongoing mediation of understanding of records (and thus phenomena), or that aspect of record making which shapes this understanding through such functions as records appraisal, processing, and description, and the implementation of processes for making records accessible."[33] This definition can easily encompass path dependence as one such phenomena which includes environments such as recordkeeping systems, both digital and non-digital, because they are the machines – both bureaucratic and physical – which guide the record making process.

Paper-based recordkeeping systems are at the core of modern office environments and synonymous with bureaucracy, government, and business. While the digital has grown in significance we still live a world where paper is of great importance. As Bearman stated, these systems serve to ensure accountability, both to the institution itself and external scrutiny. And while understanding a recordkeeping system alone is far from tantamount to understanding the organization itself it does provide researchers with (relatively) clear paths to follow. Such systems evolve over time yet still retain elements of past procedures which impact how information is displayed, transmitted, and retained. Recordkeeping systems can therefore be described as a form of high level environments governing a broad scope of functions while computing systems from mainframes to personal desktop computers are more granular instances

[32] Bearman, "Record Keeping Systems," 30.
[33] Nesmith, "Still Fuzzy," 145.

of environments. While the technical requirements and terminology may often differ the underlying principles and potential are the same; recordkeeping systems are preserved to "maintain systems that serve as the repository[34] of organizational memory of functions, structures, and events."[35] Digital environments and the capabilities and limitations they afforded to users who worked within these systems are a part of the functions and structures of more modern organizations.

Preserving recordkeeping systems is a matter of practicality, which may seem ironic considering the work that must be done by archivists to familiarise themselves with such systems for appraisal and descriptive purposes. However, it is more logical than attempting to enforce a different organizational method upon the records which will only cause confusion for users in their research later on. "Archives appraise and accession record-keeping systems, not individual records, because record-keeping systems do not just passively reflect how the creating organization used information; they actively determine it. As such, record-keeping systems are an organic whole."[36] This organic element is what allows archivists and users to glimpse into the remnants of the culture that created and used the records. Cook praises Bearman's push to place greater emphasis upon such systems in archival work: "Understand the creation and authorship of records, their animating functions and activity, their transmission and system interconnections, he advised, and the importance or value of the resulting records will almost be self-evident."[37] Expanding provenance to include a wider range of creators and the means of record production

[34] For clarification, a repository is merely a place for storage, this does not constitute an archive on its own.
[35] Bearman, "Record-Keeping Systems," 23.
[36] Ibid, 22.
[37] Terry Cook, "The Impact of David Bearman on Modern Archival Thinking: An Essay of Personal Reflection and Critique," *Archives and Museum Informatics* 11 (1997): 20.

would be ideal for electronic records in particular since it would account for multiple authors and functionality tied to specific environments.[38]

For example, the Department of the Interior was established in 1921 for the purpose of upholding the broad mandate of developing Canada's western resources "for the purposes of the Dominion."[39] The point of interest, with regards to this chapter, is the Department's record-keeping system, particularly concerning correspondence. Incoming correspondence was individually numbered and chronologically sorted in registers and then into larger dockets, indexed alphabetically by author or subject and recorded within a separate ledger, while all outgoing correspondence was copied and kept within a separate letterbook, in a fragmented system too cumbersome and complicated.[40] Developing a more coherent recordkeeping system first required consolidating the Interior and its underlying branches by separating its recordkeeping from the previous system going forward. Records were now consolidated by

[38] However, Cook has been critical of the potential limitations of Bearman's efforts to contribute to archives, particularly with regards to theory and practice. He notes that Bearman often does not consider various constraints such as budgets, legislation, demands of clients and partners, or the difficulties in training staff. Such constraints are all factors that influence preservation programs, usually as limitations, and often dictate what is practical. Bearman's ideal of reengineering modern digital recordkeeping systems to track and stamp records with what he has identified as functional requirements has the potential of bridging the gap between record-keeping and archiving. A major issue is the overt focus solely upon modern and future computing. Cook notes this lack of balance between legacy and future systems and the attitude of abandonment Bearman has exuded: "let the current stuff go and concentrate on fixing things so we won't be in the same situation ten years from now that we are in now."

This approach is problematic because it essentially proposes the abandonment of current preservation efforts and materials for the sake of a "fresh start". All efforts up until this point are rendered valueless. Going forward archivists will, somehow, have gained the authority to dictate the development of future recordkeeping systems to meet their preservation requirements. If records and systems do not survive beyond this theoretical line in the sand then they are of no lasting value. This implies that current efforts to negotiate between theoretical and practical methods of digital preservation of legacy systems/environments has nothing to teach us. There are also the moral legal concerns, as Cook points out, with abandoning systems which currently provide "significant evidence of actions, citizen's rights and status, and societal activities." While digital preservation is ever evolving it would be naïve to assume that complete abandonment of current methods, however flawed, would produce a record-keeping Eden. Instead it would create societal and informational gaps effectively creating historical "dead spots" akin to a self-imposed digital dark age. While such an approach is obviously impractical, and downright unacceptable for reasons Cook has listed and beyond, it also highlights the significance of record-keeping systems and why it is important to better understand them for the sake of more thorough digital preservation. [Cook, "The Impact of David Bearman."]

[39] Terry Cook, "Paper Trails: A Study in Northern Records and Northern Administration, 1898-1958." in *For Purposes of Dominion: Essays in Honour of Morris Zaslow,* ed. Kenneth S. Coates and William R. Morrison (York: Captus University Publications, 1989): 15.

[40] Cook, "Paper Trails," 15.

subject and listed in detailed indexes, and sectional maps were used as comprehensive guides to then relatively unsurveyed territories which detailed: active government and non-government officials; land and water routes; and travel schedules to aid in rational and informed decision making.[41]

The Department of the Interior serves as an example for how important it is to look beyond the immediate record with regards to appraisal and preservation. Viewed individually or simply by subject the records reflect little more than their individual content, but by taking a step back and applying a more critical eye it is possible to see this shift in functionality. Rather than the more passive system where each individual piece of correspondence was a treated as a single isolated issue, the Interior's new subject file system allowed for sophisticated cross-referencing.[42] This shift reflects not only the trend within the northern branches to centralize functions in response to the unwieldy volume of records being produced, but the wider movement of businesses and governments responding to the paper explosion born out of the First World War.[43]

As Bearman noted recordkeeping systems are an organic whole, and with this and Cook's examination of the Department of the Interior records in mind, we can see how environments are essential to understanding records and their creators. The shift to the subject-based system was the Department's response to the limitations of the docket-letterbook system which was too unwieldy to handle the volume of records and their interconnectedness. Therefore, it was essential to adapt by implementing a more sophisticated and centralised system that offered greater functionality by grouping related records for ease of access and cross-referencing. Modern recordkeeping systems utilize the capabilities of digital environments which are capable

[41] Ibid, 23.
[42] Ibid, 15-16.
[43] Ibid, 21.

of providing far more sophisticated functionality than traditional paper-based ones. By preserving and studying these systems it is possible for future observers to understand how creators used information.

Machine-Readable Records

Prior to the development and propagation of purely electronic records, information used for computing purposes was created and stored in paper form, specifically the punch (or punched) card. However, punch cards would come under scrutiny along with the early computers they were designed for as the move towards automation in the workplace began, scrutiny that would also encompass newer physical media in the forms of film and magnetic tape. Unlike CDs and flash memory, early forms of digital media (originally referred to as electronic media or machine-readable records) preserved much more "raw" content (no compressed MP3 files with metadata detailing the creator, length, bit rate, etc.) yet they were still reliant upon special equipment in order to access it. Archivists and other professionals, particularly of scientific background, were already trying to come to grips with the difficulties in preserving non-paper-based content.

American inventor Herman Hollerith developed some of the earliest electromechanical tabulating machines in the late 1800's which were first used to perform statistical calculations in the 1890 U.S. Census.[44] The tabulating machines ""read" or sensed holes punched in designated columns and rows on cards, and counted on the basis of the location of each hole within the column."[45] Punch cards are one of the oldest forms of digital record since they encode information using discrete variations, either there is a hole or there is not, and allowed for information to be represented semantically and therefore communicable between the fields of

[44] Margaret O'Neil Adams, "Punch Card Records: Precursors of Electronic Records," *The American Archivist* 58 no. 2 (Spring, 1995): 183.
[45] Adams, "Punch Card Records," 183.

mathematics, physics, and engineering.[46] The adoption of punch cards and early computing machines was a matter of practicality tied to the growth of states in the late nineteenth century, such as Britain, which saw governments expand to encompass more responsibilities. Barbara L. Craig notes that "The scope of its work expanded well beyond legitimate government activities of the eighteenth century, becoming, in the process, a powerful influence on the nation's social structures and in its economy. At the same time, the arrangements for the work of its offices became very much like that of large commercial and manufacturing enterprises."[47] Such growth logically meant greater workloads for departments and staff. As the government became more involved in so many aspects of the nation's life and work the consequence was more records to keep track of. Punch cards and adding machines served to augment the capabilities of institutions so that they could handle the increased demands placed upon them. "Punch card machines, tallying and sorting machines, accounting and adding machines, addressing machines, large volume duplicators, and their subsidiary pieces of equipment quickly became vital to those departments whose work could no longer be done either by hand or by traditional techniques using other devices, for example, the printing press."[48] With such expansion and adoption of new methods and equipment workplace environments began to change in order to accommodate the rapidly increasing influx of records and information that resulted from government expansion.

As with any newly adopting technologies mechanical (and later electronic) computing equipment became subject to scrutiny. Concerns of obsolescence and its impact on machine-readable records arose whenever computers were adopted and often focused not only on the records but also the dedicated equipment, the machine-readable environment. John C. Mallinson

[46] Luciano Floridi, *Information: A Very Short Introduction* (Oxford: Oxford University Press, 2010): 28-9.
[47] Barbara L. Craig, "Machines, Methods, and Modernity in the British Civil Service, c. 1870-c.1950," *Journal of the Society of Archivists* 30 no. 1 (2011): 63.
[48] Barbara L. Craig, "Machines, Methods, and Modernity," 66.

voiced such concerns in 1986, in reference to the findings in a White Paper published by Subcommittee C of the Committee on Preservation of the National Archives and Records Administration (NARA) four years prior.[49] Specifically, the issue was whether video and digital magnetic tape would be readable within one hundred years. Mallinson firmly stated that "The subcommittee slowly but surely came to the conclusion that the answer to the question was moot, or almost irrelevant. One hundred years from now, no one will know the answer because other factors are operative, the most critical of which is the short lifetime of the machine itself."[50] While digital media does have a much shorter lifespan than traditional non-digital materials (a surviving medieval period Bible made of parchment can attest to this) the Subcommittee, and by extension Mallinson, were more concerned with the machine-readable environments. Mallinson notes that the Subcommittee concluded that all content should be transferred to

> human-readable microfilm, which guarantees indefinite preservation, in the old-fashioned human-readable sense and not in the modern error detection, correct, and re-recorded sense. It provides a hardware and software independent mode of preservation. It is amenable to mechanization, that is it can be surrounded with all the glories of technology.[51]

The issue with such a method of preservation is that is essentially isolates the record away from the "glories of technology" that enabled their creation. For the purposes of quick reference and security in case of damage to or loss of the necessary environments this is a useful method of preservation, but on its own it severely curtails any potential future use by removing the functionality of the records.

Canadian archivist Sue Gavrel had concerns which extended beyond simply preserving information. They lay with how to also preserve the usability of the information. Electronic Data Processing or EPD records, as she refers to them, are valuable because the ability to manipulate

[49] John C. Mallinson, "Preserving Machine-Readable Archival Records for the Millenia," *Archivaria* 22 (1986): 147.
[50] Mallinson, "Preserving Machine-Readable Archival Records for the Millenia," 148.
[51] Ibid, 152.

them allows for repeated re-analysis of non-aggregated data for statistical purposes. Keeping large amounts of data on microfilm limited users and researchers to traditional methods of calculation. This would require recopying and recalculating everything by hand or having to reproduce the data in a machine-readable format every time someone wanted to use it. This also runs counter to Mallinson's concerns that the cost and manpower required to re-record digital information to newer formats would be too prohibitive to even contemplate.[52] Gavrel argued that "Conservation is an active process, requiring the archivist to keep abreast of changes in technology. Tapes cannot be stored on a shelf and forgotten" and that "Although the medium is vulnerable, the recopying of data ensures its long-term preservation. Although the value of the records is linked to their manipulability, there is no intrinsic value in the medium."[53]

Becker somewhat echoes this sentiment with his assertion that preservation is about what constitutes an authentic reproduction of a record, or performance. Gavrel argues that so long as the data is in a form that is still machine readable, even if it is on different media, then the data is still properly preserved. However, she acknowledges that despite the limitations of obsolescence archivists should embrace digital technologies and that "The more tools technology has provided, the more inventive researchers have become in using resources."[54] Mallinson argues for media migration in addition to format migration. Some archives preferred printing out born-digital records assuming paper was more authentic, since it lacked the "disturbing ephemerality", or inherent manipulability of the digital. Mallinson's proposed re-formatting from manipulable digital data to fixed microfilmed data represents a similar shift.[55] Gavrel, however, followed a more modern approach arguing for a more active role in preservation through understanding the

[52] Ibid, 151-2.
[53] Sue Gavrel "Preserving Machine-Readable Archival Records: A Reply to John Mallinson," *Archivaria* 22 (1986): 154.
[54] Gavrel "Preserving Machine-Readable Archival Records," 155.
[55] Greg Bak, "How Soon Is Now? Writings on Digital Archiving in Canada from the 1980s to 2011," *The American Archivist* 79 no. 2 (2016): 296.

changes in technology and moving contents forward to preserve manipulability, which some archivists perceived as "an important aspect of their digitality"[56] and what makes them so invaluable.

In 1984 a UNESCO Records and Archive Management Programme (RAMP) study was presented in Paris by Harold Naugler, a machine readable records archivist at PAC. Its purpose was "to provide archivists and other interested information professionals with an introduction to machine-readable records and to provide guidelines for the appraisal of their archival value."[57] The RAMP study offered a lengthy examination of computer hardware, software, and storage mediums, as well as their functions and limitations. In his findings, Naugler stated that, by their nature, the information on machine-readable records "is accessible, interpretable, manipulable, and transmittable only by automated or electronic means" and should be thought of as "a dynamic entity having certain organic properties being composed of unique, fundamental, and discrete bits of information or data elements which can be rearranged, changed, manipulated, merged, or deleted in order to generate a set of information on demand."[58] Gavrel and Naugler both identified the dynamic or organic nature of information stored on machine-readable records as more than just a feature. It was the defining aspect and just as important to include in the preservation process as the content. Naugler specifically pointed out that

> It is not sufficient to have access to the machine-readable information or raw data themselves. In addition to the data, it is necessary for an archivist to obtain documentation which describes the contents, arrangement, codes, and technical characteristics of the machine-readable data file. Without such documentation an archivist is unable to appraise the value of the information, and a researcher is unable to access the information.[59]

[56] Bak, "How Soon Is Now?," 296.
[57] Harold Naugler, *The Archival Appraisal of Machine-Readable Records: A RAMP Study with Guidelines* (Paris: United Nations Educational, Scientific and Cultural Organization, 1984): iii.
[58] Naugler, *The Archival Appraisal of Machine-Readable Records*, 14.
[59] Ibid, 15.

As Gavrel argues, ensuring data remained machine-readable so that functionality could be reproduced would be essential for preserving machine-readable records. As the importance and volume of machine-readable records increased so did concerns and discussion over their nature and how to determine their archival value. However, the hurdle that early computers had to overcome before such concerns and discussion could arise was acceptance in both the public and working spheres.

Early Computer Adoption – Rhetoric

Computers were not an overnight success within the working and public spheres. Even as their presence and importance grew, they (and early digital records) were often seen as mere extensions of pre-existing technologies rather than unique equipment. As Mahoney stated, the computer was what we made of it, so it is only logical that this would extend to the digital record. Therefore, as I have argued, if we understand the environment, we can better understand the record. Like any other invention, computers were not developed in a vacuum nor do they operate in one. And despite the promised and actual impact of automation, human interaction is required to achieve their true potential. Archivist Jay Atherton proclaimed almost 50 years ago to

> Let it be stressed immediately, however, that a computer does not think through an intuitive process as we do. All it really does is react to a given set of circumstances according to a given set of instructions (its programme). Its degree of competence is dependent upon the accuracy and thoroughness of the programmer and the programme.[60]

This interaction between man and machine is the social component of computing. It is still one of the most influential forces in the development of computers. Atherton argued that automation is a means to extend the user's faculties with the computer performing calculations at the behest of the user.[61] Human agency has been one of the most significant driving forces in computer

[60] Jay Atherton, "Automation and the Dignity of the Archivist," *The Canadian Archivist* 2, no. 1 (1970): 57.
[61] Atherton, "Automation and the Dignity of the Archivist," 58.

development and adoption. Whether or not people perceived computers to be useful (or detrimental) to their work and their lives has impacted where, when, and how computers became integrated into society, and the degree of success.

One way of understanding the importance of the computer in society is to look at how they were advertised. William Aspray and Donald deB. Beaver identified four important factors: first was that advertising is one of the principle methods of disseminating new technologies to customers and the public; second, advertising shows us how the computer was presented to different potential users; third, advertising is used to identify and display issues that may be of interest to customers; and fourth, advertising has proven to be valuable source material in other areas of study.[62] Advertising has always been about sending a message, often by identifying problems (real or manufactured) and providing the solution, and frequently in a way that is meant to demonstrate "empowerment" or progress. Advertising for current computers and applications declaring this year's (or month's) latest-and-greatest product are little different from the ones that were used to sell mainframes.

In the 1950s the computer was, as Thomas Haigh noted, a form of symbolic modernity that drew the attention of visitors and the local news, with some businesses convinced that the publicity alone improved both public and business relations.[63] There were of course positive reactions to the arrival of the computer which fell in line with their depiction as a form of "symbolic modernity" and some went even farther, tying it in with national identity. Jaakko Suominen and Jussi Parikka identify the start of the Computer Age in Finland as 17 October 1958 when Minister of Financial Affairs Päiviö Hetemäki pressed a button to start Finland's first

[62] William Aspray and Donald deB. Beaver, "Marketing the Monster: Advertising Computer Technology," *Annals of the History of Computing* 8 no. 1 (1986): 127-9.
[63] Thomas Haigh, "The Chromium-Plated Tabulator: Institutionalizing an Electronic Revolution, 1954–1958," *IEEE Annals of the History of Computing* (2001): 87.

computer, and IBM 650 dubbed ENSI ("First"), at the Post-Savings Bank in Helsinki.[64] The computer has been advertised, in whatever physical form, as a thing of awe, and media (such as advertising) has continued to strike to evoke an emotional response to technology. Mainframe computers were anthropomorphised as electronic brains, which meant that their imposing size symbolised power – the opposite of today – and even intelligence.[65] Between 1958 and 1960 Finland rented three mainframes and constructed a fourth, dubbed ESKO, all launched to fanfare and speeches including participation from the "guests of honor" in the form of performing calculations, printing pictures, and playing music.[66]

Automation was directly associated with the factory assembly line which was held as the standard of modern efficiency. In the 1930s administrators began to adapt the office space to accommodate tabulators as part of the shift towards administrative efficiency,[67] a movement which would encompass early digital computing. Digital computers represented an evolution of the ideals and concerns surrounding automation in the workplace and society. Various different groups or "communities of computing"[68], all viewed early computers as means to different ends. Scientists and engineers strove to develop systems capable of handling logical functions beyond basic arithmetic, while business, industry and government sectors sought greater control of management, production, and communications through "tactical analysis".[69] As computers developed to become more protean in nature it meant that (theoretically/optimistically) they could serve whatever purpose these communities of computing deemed necessary. This is also reflected in the displays of symbolic modernity put on by Finland and others which were an

[64] Jaakko Suominen and Jussi Parikka, "Sublimated Attractions – The Introduction of Early Computers in Finland in the Late 1950s as a Mediated Experience," *Media History* 16 no. 3 (2010): 319.
[65] Suominen and Parikka, "Sublimated Attractions," 324.
[66] Ibid, 327-8.
[67] Craig, "Machines, Methods, and Modernity," 70-1.
[68] Mahoney, "Histories of Computing," 124.
[69] Ibid, 123-5.

evolution of society's fascination with computing and its potential. Steven Lubar points to news media such as the New York Times which "ran a story in 1940 about crowds that gathered in front of an office-supply store in Albany to watch punch card sorting machines in action", and when the "Saturday Evening Post referred to the Los Angeles Police Department's Hollerith machine as 'a mechanical Sherlock Holmes,' a 'crime-hating robot,' 'The Detective Who Never Sleeps'."[70]

However, the emotional and critical response to the expanding adoption of such automation was not always positive. Many students at Berkley University in the 1960s felt that punch cards used for registration turned them into numbers in a bureaucratic machine rather than people. Driven by this feeling of dehumanisation some went so far as to reprogram the IBM machines to print words and phrases ranging from "FSM" (Free Speech movement) and "STRIKE" to inserting obscenities into class lists.[71] Punch cards were visible, portable symbols of protest made possible by understanding and harnessing the (literal) machine that produced them. This specific form of protest required knowledge of how the environments worked in order to reprogram them to produce the slogans and obscenities used to express their discontent. As Atherton asserted, the computer performs at the behest of its user, and in the case of Berkley this meant both the administration and (occasionally) the students. Lubar notes that "The technical prank – or "hack" as it's known at engineering schools – generally serves to reinforce the importance of technology rather than to subvert it."[72] The historical significance of environments, digital and non-digital, is the sum of both their social and technical processes. To Berkley staff they were tools for automating mundane tasks for the sake of efficiency. For many

[70] Steven Lubar, ""Do Not Fold, Spindle or Mutilate": A Cultural History of the Punch Card," *Journal of American Culture* 15 (1992): 44.
[71] Lubar ""Do Not Fold, Spindle or Mutilate"," 46-48.
[72] Ibid, 54.

students they served a dual purpose: both as the source of their ire for being seemingly reduced to numbers and a tool of protest.

In order to understand the impact of environments it is important to look beyond the rhetoric and ceremony that surrounded tabulators and the electronic computer systems that followed. Mahoney wisely wrote that "hype hides history"[73] and Canadian archivist Hugh Taylor identified that with new forms of media come new ways to represent information. This means that conventional ways of thinking about what a "file" is will also change.[74] Such change must also accommodate environments since they are the point of contact between the users and the records. The electronic computer, specifically the mainframes of the 1950s, were heralded as the next industrial revolution but the reality was, they usually supplemented existing punch card systems due in no small part to the fact that they often relied upon them for input and output.[75] The potential benefits of electronic computing environments were well advertised in corporate and sales rhetoric but in practice there was no overnight revolution. In fact, computers were largely integrated into existing office workflow environments as an extension of existing tools and methods[76] while reinforcing the concept of the corporate ladder as a means of advancement for staff rather than replacing them.[77] Achieving "symbolic modernity" required a period of growing pains for customers and staff, both for machine-readable records and later electronic computers.

Early Computer Adoption – Reality

Adopting new technology will inevitably lead to a learning curve when it comes to integrating it into the workplace. While electronic computers were touted as the next industrial

[73] Mahoney, "The Histories of Computing(s)," 120.
[74] Hugh Taylor, "Transformation in the Archives: Technological Adjustment or Paradigm Shift?," *Archivaria* 25 (1987-88): 18-20.
[75] Haigh, "The Chromium-Plated Tabulator," 85.
[76] Ibid, 85-6.
[77] Ibid, 94.

revolution the fact remained that workplaces and employees do not change overnight. Though mainframes and digital computing would prove to be the future, the issue was exactly what role they would play and just how significant it would be. Most workplaces had only just gotten comfortable using punch cards and tabulators post-war yet debate still occurred concerning the historical value of machine-readable records. While tabulators were widely adopted by the 1960s and electronic computers were increasingly common, from a preservation standpoint they were still an unknown quantity.

The long-term value of machine-readable records was a point of debate in the U.S. as the government and National Archives tried to come to a consensus over whether punch cards were worthy of preservation. In the 1930s the Accessions Advisory Committee stated the government agencies which created and used punch cards were responsible for determining whether they held any value, not professional archivists.[78] During the 1950s deputy examiner T. R. Schellenberg, who would emerge in the 1960s as a leading archival thinker, argued that physical form was a determinant of the value of a record and that records should be accessible without requiring any sort of equipment, and if the agency that produced records "did not fully exploit them" it was unlikely anyone else would have the means to do so themselves.[79] At the time it would have been difficult for users to access the equipment necessary to utilize punch cards so concerns over their archival value are understandable. However, this argument fails to account for the functionality of the punch cards and the potential value of the data for future research Gavrel was concerned about.

The general attitude of machine-readable records not being equal to other paper records, and the removal of archival input in the appraisal process led to what Margaret O'Neil Adams

[78] Adams, "Punch Card Records," 193.
[79] Ibid, 196.

describes as a "tone of disinterest" that persisted into the era of magnetic tapes.[80] This was all the more unfortunate because despite the fact that, while electronic computers replaced mechanical tabulators, they still initially utilised punch cards as a medium for inputting information which could now be saved onto electronic memory for more rapid access and greater range of analysis.[81] Despite eventually being rendered obsolete by magnetic media, punch cards were able to outlive the mechanical environments which created them. This attitude was even more disconcerting considering the Roper Center, privately founded a decade before, was collecting punch cards from as far back as 1936. Other independent efforts were soon to come, embracing new technologies and adopting early media migration from card to magnetic tape.[82] The National Archives was being set to fall behind the technological curve and unprepared for the rapid evolution of computing environments.

For staff at the Public Archives of Canada (PAC) the 1960s was a period of growing pains as archivists worked to familiarise themselves with the intimidating and unfamiliar jargon and procedures of the computer age, while grappling with concerns over whether the arrival of the computer would undermine their profession and expertise, or if it was a means to higher standards of records management.[83] The Glassco Commission Report, focused on Paperwork and Systems Management within the public service, stated that "There is urgent need for a comprehensive plan to control the products pouring from typewriters, duplicating machines, and high speed printers of electronic computers."[84] A glaring oversight found by the report was that "Departments have no real guidance on how to conduct the preliminary studies that must precede

[80] Ibid, 197-8.
[81] Ibid, 186.
[82] Ibid, 196-7.
[83] Betsey Baldwin, "Confronting Computers: Debates about Computers at the Public Archives of Canada during the 1960s," *Archivaria* 62 (Fall 2006):159-60.
[84] Canada, Privy Council office, "Report 4: Paperwork and Systems Management," in *The Royal Commission on Government Organization. Vol. 1 : Management of the public service / J. Grant Glassco, chairman* (Ottawa, ON): Queen's Printer 1962: 486. http://publications.gc.ca/site/eng/471934/publication.html.

any intelligent decision to use automatic data processing methods."[85] Betsy Baldwin also notes that the report failed to consider magnetic tape-based media, despite the work of the Commission's own ADP (Automatic Data Processing) subgroup analysing and reviewing active federal computer installations, and did not even include any recommendations concerning the preservation of computer records with regards to the Public Archives.[86]

The adoption of tabulators and computers did not herald a total revolution for government records production and management. The initial response was sporadic at best. This is due in part to the fact that neither computers nor computer records had entered into use in most archives, so the lack of familiarity and formal appraisal methodology was arguably a matter of a lack of any precedent.[87] The exclusion of archivists from formal decision making concerning computer records also meant that there was little formal need for archivists to invest time and effort familiarising themselves with the records or environments. Greg Bak identifies that the lack of exposure to computers influenced the opinion Canadian archivists had for the computer "revolution:"

> Computer records were viewed as a niche preoccupation, a form of media specialization. Like other media specialists, digital archivists of this era – specialists in what was then called machine-readable archives (MRA) – wrote and spoke about their work in relation to mainstream textual archives, highlighting points of similarity and contrast. Perhaps as a result, their writings were perceived to be secondary to and derivative of developments within textual archiving, rather than a distinct tradition of their own.[88]

Despite the influence of computing in businesses and government, computers and electronic records were treated as existing "outside" the regular working environment. Dominion Archivist W. Kaye Lamb emphasised in 1968 that any records related to computers, electronic or

[85] Canada, Privy Council office, "Report 4: Paperwork and Systems Management," 497.
[86] Betsey Baldwin, "Stepping off the Paper Trail? Rethinking the Mainframe Era at the Public Archives of Canada," (PhD diss., University of Ottawa, 2006): 72-6.
[87] Baldwin, "Stepping off the Paper Trail?," 79.
[88] Greg Bak, "Media and the Messengers: Writings on Digital Archiving in Canada from the 1960s to the 1980s," *Archivaria* 82 (2016): 56.

otherwise, were transitory or "working papers" and did not need to be archived.[89] By 1967 PAC was "among only 21 of 48 departments that had no computers or terminals installed on site."[90] PAC was now in a similar position to what the U.S. National Archives was experiencing concerning electronic records. It was also ill-equipped to handle the increasing reliance on computing environments and the unique challenges they presented with regards to preservation.

In 1973 after sending archivists to study the methods of the US National Archives and Records Service (US NARS) a Machine Readable Archives Division (MRA) was established at PAC, yet was severely limited by the fact that there were no computers in the reading room. Users were limited to an "elite" who could afford access to computers beyond the reading room.[91] There was a shift towards greater appreciation for the digital record as being unique and that the appraisal process had to focus on more than just the record itself. MRA writers "emphasized that digital records are distinctive and require their own processes, but that their acquisition should be fully integrated with other media, reflecting the broader goals of the appraisal program and in keeping with the resources of the institution."[92] Part of this shift was the fact that computing environments were no longer solely the tools of government and big business. Rather than requiring an entire department devoted to managing a large mainframe, computers had begun to shrink and were advertised with individual offices as their target markets.[93] The combination of diminishing size and increasing power meant that computing environments became more widely accessible and began to enter the public sphere. The once massive "electronic brains" were slowly becoming regular fixtures of everyday life.

[89] Baldwin, "Confronting Computers," 176.
[90] Baldwin, "Stepping off the paper trail?," 83.
[91] Bak, "Media and the Messengers," 62-3.
[92] Ibid, 64.
[93] Aspray and deB. Beaver, "Marketing the Monster,"131.

The 1980s and 1990s saw a major boom in the presence of computers in both the workspace and eventually the home. This led to a shift towards desktop and personal computing, both of which heralded significant changes in the work and personal spheres. Previously, users had to schedule access to a limited number of mainframes or minicomputers, often shared by several departments. As computers grew smaller individual users could now have an entire computing system on their desks, both at work and in the home, making modern computing accessible to the individual. Computing was becoming more personal and was drawing attention from hobbyists and tech-savvy archivists. Unfortunately, it also led to a knowledge and culture gap between long-term users of computers and newer users for whom personal computing was their first proper introduction to the digital. The work of the MRA staff was seemingly ignored or dismissed as Canadian archivists instead turned to specialists or "systems gurus" for guidance rather than the established work of their MRA colleagues.[94] The arrival of electronic computing in the form of the mainframe – touted as an actual "revolution" – was primarily a means of advancement to which punch card staff aspired and adapted their current skillsets. To the previously uninitiated it was also difficult to connect with MRA archivists and others who had to work with computers of the 1960s and 1970s; the development of ready-to-install software and point-and-click graphical user interfaces (GUIs) made it appear relatively easy to enter into digital preservation and many saw the efforts of the MRA as overly technical rather than truly archival.[95] This increased sophistication in the capabilities of digital environments to manipulate information and records brought about renewed discussion over how best to preserve them. Rather than attempt to take advantage of the improved user-friendliness of GUIs and third-party

[94] Bak, "How Soon Is Now," 287-8.
[95] Bak, "How Soon Is Now?," 291.

software archivists instead worked to reinforce the methods of traditional paper-based environments.

In what was arguably a step backwards, most archival institutions chose to make no changes to their current methodology to accommodate digital records, instead simply printing them out and preserving the paper copies rather than the original format, possibly out of a sense that they were more "authentic" since they could no longer be edited without leaving evidence of such.[96] In doing so a blind eye was being turned towards to the significance of digital environments and declaring they produced "non-records", essentially a continuation of Dominion Archivist W. Kaye Lamb's assertion they were transitory "working papers". The inherent manipulability of digital records was meant to serve a limited purpose before the final output was "set": "Much like the "old-timers" who print and file their email or PowerPoint™ slides from presentations attended, late twentieth-century users considered the computer a transitory tool designed to move information from mind to paper."[97]

In the 1990s, another approach came from the more recent development of the Information Management and Office Systems Advancement (IMOSA), built upon the MRA focus that early intervention was important in preserving digital records. It differed in approach from its predecessor in that it "sought not to understand and document the digital technologies in use by a records creator; rather, IMOSA sought to create an artificial environment to replace the digital technologies in use by records creators."[98] Rather than integrate computing environments, like businesses and government employees had done previously in response to mainframes, archivists were developing policies based upon working *around* them. The ability to rapidly

[96] Ibid, 296.
[97] 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): 242.
[98] Bak, "How Soon Is Now?," 297.

manipulate records and their contents was the cornerstone upon which the expectations of computers had been built. Records were becoming more dynamic thanks to computing environments, yet archival policy did not reflect this. Instead efforts were made to maintain the status quo of a record being a definitive work outside the purported realm of "transience" described by archivists such as Lamb. The 1980s and 1990s saw a shift in how archivists approached computers and preservation concerning digital records. It was assumed "the solution to digital recordkeeping challenges lay in getting computer developers to better understand the importance and nature of recordkeeping requirements: that it was up to the computer scientists and systems designers to create products that would conform to recordkeeping requirements."[99] This distancing did nothing to stem the tide of digital records entering archives and archivists struggled to find a balance between theory and practice, and how best to preserve digital records and the systems that they originated from.[100]

In response to the increasing volume of records entering archives a number of tools and methods have emerged with the purpose of addressing long-term preservation and access. These have often been the result of collaboration between archives, computer scientists, and other concerned parties within the field of digital preservation. Examples such as the development of the Open Archival Information System (OAIS), and works such as the PREservation Metadata: Implementation Strategies (PREMIS) data dictionary, intended to support OAIS implementation, have "embraced participation from archivists, information and communications technology professionals, and others."[101] Despite shifting assumptions archivists have continued to embrace the technical resources and know-how to preserve digital materials. This includes the adoption and integration of legacy computing systems and components as well as digital forensics tools

[99] Ibid, 297-8.
[100] Ibid, 300-1.
[101] Ibid, 304.

into preservation workflows and contributing to the ongoing development of software preservation systems such as Archivematica. Emulation has already seen use in individual projects while recent efforts to build upon cloud-based service models may prove to be the next big step in enabling users to experience original records and environments in a more immersive way. Reference models and descriptive standards have continued to evolve to better incorporate digital record creation, preservation, and infrastructure. Chapter Two focuses upon these tools and methods and how they can be used to better leverage the potential of digital records to better express the record-environment relationship to users.

Chapter Two - Methods of Preservation

Where Are We Now?

Digital preservation has become increasingly important as our dependence on digital technologies has grown over the decades, particularly from the 1990s onward with the rise in use of the internet for business and personal use and the impact the smartphone has had in both the work and social spheres. This rapid march forward does not mean that we are completely free of reliance upon obsolete equipment and record formats, nor should we dismiss their impact, as covered in the previous chapter – quite the opposite. Governments and businesses often still rely on legacy applications alongside newer software since much of their operations are built around them, while valuable historical information resides on obsolete media languishing in storage and caught up in archival backlogs. Fortunately, a growing emphasis placed upon digital preservation and growing sophistication in tools available has given rise to several different methods of preservation, the four major ones being: format migration, the most widely adopted; preservation of physical hardware, often identified as legacy or retrocomputing; the use of emulation to simulate older environments on newer hardware; and descriptive and other metadata standards, combined with other documentation practices to help identify and manage information for the purposes of digital preservation or even in lieu of preservation of the original data or application. Each method has its strengths and weaknesses regarding the authenticity of the records or information being preserved (both reproducing the record and the experience of interacting with the environment); accessibility (ease with which users may access and interact with the environment and in turn records; and resources (time, money, expertise, and equipment).

The provenance of a digital record or object, including environments, is a valuable source of information regarding the basic requirements for preserving digital content. Terry Eastwood

points out that by the end of the 1980s it had become impossible to preserve many electronic records and legacy systems as "little attention has been paid to identify important facts of their creation, relationships, and context, including their technological context."[1] As discussed in Chapter One the tools and technologies involved in the development of digital records, in particular the digital environments, are part of the history of the record and the point of contact between the record and the user. Eastwood argues for change, starting with the appraisal of digital records or objects, noting the importance of compiling information about them since "it is probably a mistake to think that data on its own will be sufficient for future understanding."[2] He explains that within this collected documentation archivists will find greater context for the records including "the provenancial context, [which] refers to the creating body, its mandate, structure, and functions."[3] Terry Cook noted a shift away from focusing solely on content to "maximizing the power of creating and using **records in context**, that is, provenance."[4] As Eastwood asserted, content on its own has limited meaning which may diminish over time. This is because later users will be increasingly removed from the context in which it was created and encounter ever growing knowledge gaps. Cook summarises the role of the archivist and significance of provenance:

> archivists want to know (and share with their researchers) not just what was communicated, but when, by whom, to whom, where, how, why, in relation to what other documents, using what media, connected to what broader functions, programmes, and activities, both now and over time. These broad questions rest on three properties which all records have: content, structure, and context.[5]

---

[1] Terry Eastwood, "Appraising Digital Records for Long-Term Preservation," *Data Science Journal* 3 no. 20 (December 2004): 203.
[2] Eastwood, "Appraising Digital Records," 204.
[3] Eastwood, 204.
[4]Terry Cook, "Electronic Records, Paper Minds: The Revolution in Information Management and Archives in the Post-Custodial and Post-Modernist Era." *Archives and Manuscripts* 22, no. 2 (November 1994): 300–29. Reprinted in *Archives and Social Studies: A Journal of Interdisciplinary Research* (March 2007): 405–49. Emphasis as in original.
[5] Cook, "Electronic Records, Paper Minds," 426.

In short, digital preservation is about more than just the record, it encompasses its provenance, purpose, and the physical and theoretical processes that made it and give it identity and purpose. Digital environments are an essential part of the creation context because without them there is no digital record. However, being able to create and use records in context has limits. When well preserved, content can still be read and understood in a coherent manner, even if separated from the context of creation. So long as the content can be reproduced – such as through format migration – so that it may be interpreted it is of potential value to future users. My point is to look beyond the content to reconsider the technical context of its creation – in this case, environments – and how they may also be preserved in a manner that makes them as accessible to users as the content.

This chapter is divided into four sections, each devoted to examining the methods of preservation mentioned above: format migration, legacy computing, emulation, and description and documentation. This will be done by drawing upon recent projects – completed and ongoing – ranging from hobbyist movements to institutional and international efforts dedicated to tackling the issues of digital preservation. This chapter will also examine how (or if) digital environments fit into these efforts and how effective they are. It is important to note that there is no inherent mutual exclusivity to any of the four methods of digital preservation discussed in this thesis. They can be, and, as will be noted in the sections below, are often used together to varying degrees.

Methods of Preservation 1: Format Management and Format Migration

As discussed in Chapter One format migration is currently the predominant form of digital preservation and is also preventative in that it is a means to remain a step ahead of obsolescence. While methods such as emulation are focused upon reproducing older systems

"[format] Migration, on the other hand, aims at the integration of past content into future knowledge-based workflows."[6] The goal is to make records accessible by transforming them from at-risk formats into more open or widely adopted ones and therefore making them less dependent upon specific hardware and software. Risks which may occur during the format migration process or subsequent to it (such as data loss or mismatch of certain functions) can be identified either through supporting documentation or through dedicated format migration tools but this does not completely eliminate the possibility of error. Chris Frisz et al. point out that "The primary sources of these risks are mismatches between source and target formats and differences in interpretation" which can sometimes occur as formats diverge from one version to the next.[7] Other issues such as digital decay (or bit rot) and the possibility of viruses, among others, mean that digital records can be very "hands-on" with regards to preservation. If no sufficient alternative format can be identified then certain records may not be submitted to format migration out of concern for such differences in interpretation.

Identifying suitable format migration pathways can be performed manually or automated through dedicated software. Archivematica is a free, open source, and standards-based long-term digital preservation system.[8] Among the numerous functions it can perform include identifying formats, generating preservation metadata following international standards, and performing format migration, all of which can be automated.[9] Digital records require continuous monitoring from the point of ingest when they first enter into archival holdings through to the point where users can readily access them. Archivematica was developed by Vancouver-based Artefactual

[6] Christoph Schlieder, "Digital Heritage: Semantic Challenges of Long-term Preservation," *Semantic Web* 1 no.1 (2010): 144-45.
[7] Chris Frisz, Geoffrey Brown and Samuel Waggoner, "Assessing Migration Risk for Scientific Data Formats," The International Journal of Digital Curation 7 no. 1 (2012): 30-6.
[8] Bronwen Sprout and Mark Jordan, "Archivematica As a Service: COPPUL's Shared Digital Preservation Platform," *Canadian Journal of Information and Library Science* 39 no. 2 (2015): 236.
[9] Sprout and Jordan, "Archivematica As a Service," 240.

Systems Inc., including initial funding from UNESCO, and in collaboration with the City of Vancouver Archives and the International Council of Archives (ICA). The purpose behind this effort was "the aggregation and development of an open source archival system, building on and drawing together existing open source programs."[10] Archivematica is a system that integrates multiple open source tools in order to accomplish these tasks. The free to use and open source approach means that it is available to anyone and any institution, regardless of size, to help manage their digital records.

Unfortunately, such systems and tools are not always readily available to handle certain document types which can lead to the scenario in which the "formatting of the converted document often differs from the original; the differences may be minor but they can change the overall look and feel of the document, which may call into question the authenticity of the conversion."[11] This can force archivists to adopt a wait-and-see approach, such as the one McLellan outlines: "As with moving image files, acceptable conversions might not be possible for the immediate future and bulk Microsoft Office migration processes may need to be run at a later time when better tools become available."[12] However, Archivematica, and similar software-based systems, are not complete digital preservation programs in and of themselves. While users can automate specific tasks such as the capture and generation of metadata, Archivematica is geared towards preservation, not archival description. It cannot identify levels of description (fonds, series, file, and item, for example) nor potential categories of materials (correspondence, accounting, or blueprints, for example) upon ingest. This very much requires manual input and insight from living, breathing archivists. Their purpose is to address specific technical issues and

[10] Evelyn Peters McLellan, "Selecting Formats for Digital Preservation: Lessons Learned during the Archivematica Project," *Information Standards Quarterly* 22 no. 2 (2010): 30.
[11] McLellan, "Selecting Formats for Digital Preservation," 32.
[12] Ibid, 32.

automate procedures, particularly in cases where digital records are handled in bulk, where possible. While Archivematica "supports emulation preservation plans by preserving original bitstreams"[13] it does not ensure this functionality as Archivematica is not itself an emulator.

All preservation is an ongoing process, but it takes on new degrees of complexity with regards to digital records. Existing processes do not lend themselves well to preserving digital environments. It is possible to utilise format migration to make some records more widely accessible by utilising formats that make them relatively environment-independent, and it can be done easily with dedicated tools, many of which are free. This means that obsolescence is not automatically a death knell for data in many older records. From a custodial perspective the consequence is that, unlike paper records, which can be placed in a climate-controlled room and left untouched for years, digital records require regular review to ensure they remain up to date. From a functionality perspective there are limitations and trade-offs that may include loss of features or even data either due to the format migration process or the properties of the newer formats and applications. This is why original versions of records are always preserved, even if they are not intended for public viewing. Unfortunately, environments cannot be made compatible with newer systems at the click of a button since they are far more complex than the records created and manipulated through them. Therefore, transformation efforts are focused primarily upon records since they contain the content that the majority of users will desire access to. The issue is then communicating to them that one or more transformations have taken place in order to make the content accessible in an (ideally) relatively consistent manner, and how they view them now, within the context of their modern computer system, differs from the original format and system.

[13] Ibid, 30.

Methods of Preservation 2: Preservation of Older Hardware and Software

How the Obsolete is Still Vital

Preserving older hardware and software is the most direct way of maintaining access to older digital environments and digital records since it does not require migrating information to alternate formats or utilizing emulation to reproduce select properties or obsolete (or legacy) systems. The most significant and recognisable benefit of this method is that it allows users to experience the record in its original form and experience the same capabilities and limitations as the record's creator. However, simply having the technology available does not instantaneously grant accessibility; that requires the necessary technical knowledge and experience. The sheer number of unique and derivative operating systems developed throughout the 1970s, 80s, and 90s would perplex users who have little familiarity with computers beyond modern Windows or Apple products. They are unlikely to intuit how to interact with a pre-graphical user interface, such as the command line disk operating system (DOS) on legacy Atari, Commodore, or BBC microcomputers, for example. Rather than the simple point, click, and drag to access and move files and folders, many legacy systems require users to directly type in commands including the directory and file name (including extension) to access or move contents around.[14]

Need is the most significant prompt for identifying what older hardware must be sought out and preserved. Need can potentially be determined simply by looking through backlogs and archival holdings to identify the forms of digital media an archive has in its possession. If there is no description of the contents stored on a form of digital media, or said contents are known and considered potentially worthy of preservation, then the only way to access it is to render it. It is usually while in either of these two scenarios that archivists experience the impact of inaccessibility brought about by the lack of access to compatible environments. Therefore, one of

[14] "How to Use MS DOS," wikiHow, last updated August 8, 2019, https://www.wikihow.com/Use-MS-DOS

the few options available to access the content is to acquire functional hardware and software components of an obsolete environment.

In 1999 the Archives of Ontario began a legacy electronic media pilot project that identified over two thousand instances of obsolete portable digital media in their holdings, much of which contained content that was government produced.[15] Concerns regarding the fragility of magnetic media (similar to those expressed by Mallinson in the previous chapter) led to the adoption of a policy where "the focus has been on information preservation, not on media preservation."[16] While Mallinson's concerns over long-term media stability, and Gavrel's rebuttal that functionality is key, represent a major dichotomy in digital preservation theory and practice, the argument is rendered moot if one cannot even read the media to begin with. This was one of the issues faced by the Archives of Ontario when it came to the daunting task of accounting for the electronic media in its holdings, including 5.25-inch floppy disks. Finding hardware that could read the disks was difficult: even a local company called FreeGeek (which specialized in rebuilding and recycling older systems at no cost other than supplies and goodwill) had no compatible drives. It was fortuitous that another department had built, in-house, a compatible machine made by a staff member from parts found in a basement.[17]

One challenging aspect of preserving older digital environments is actually getting hold of them, and necessitated in the Archives of Ontario looking elsewhere for a solution. Charles Levi noted the implications of the gradual disappearance of obsolete technologies and that archivists must invest in the past in order to handle such technologies in the future: "Archival repositories need a commercial operator/developer willing to return to the business of producing 5.25-inch

[15] 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): 240-1.
[16] Levi, "Five Hundred 5.25-Inch Discs," 241.
[17] Ibid, 243.

drives. Such an endeavour would, however, require a profitable market, which is unlikely. Failing that, a consortium of archives could sponsor an archival 5.25-inch construction project. The schematics and materials must still exist somewhere."[18] Obtaining a drive is also not an immediate solution since it must be connected to a compatible computer system. Fortunately this does not necessarily require an entire legacy systems as there are modern tools that help bridge the gap between generations of computing systems.

The KryoFlux Preservation Technology Group (KFPTG or simply KryoFlux from hereon) partnered with The Software Preservation Society, and Digital Intelligence have produced modern hardware and software solutions to interfacing with and obtaining data from 3.5-inch, 5.25-inch, and even 8-inch floppy disks. KryoFlux and the Software Preservation Society have developed a unique piece of hardware (also called KryoFlux) that connects to a floppy drive utilizing a traditional 34 pin connector while a USB 2.0 connector is used to connect to the computer.[19] Such simplicity is more impressive when compared to the situation faced by the Archives of Ontario, which required building an entire computer system in order to accomplish the same task. While users will still require a working floppy drive to actually read the disks, the software and drivers included with the KryoFlux device support modern Windows, Mac, and Linux systems and enable users to create disk images[20] in formats supported by numerous older DOS including Commodore, Atari, and others.[21] KryoFlux essentially acts as a bridge between legacy peripherals and modern digital environments by streamlining the process by which data is accessed and copied for preservation purposes. It allows data to exist on modern

[18] Ibid, 244.
[19] "Technical Overview," KryoFlux, Accessed October 23, 2018, https://www.kryoflux.com/?page=kf_tech.
[20] A disk image is an exact bit-for-bit copy of the contents of a data storage device (including files, software, operating system, etc.). This could be a CD, flash drive, or the contents of an entire computer hard drive.
[21] "Features," KryoFlux, Accessed October 23, 2018, https://www.kryoflux.com/?page=kf_features.

systems, which increases accessibility without the need to repeatedly handle delicate and at-risk legacy media and environments.

KryoFlux has seen success as a powerful tool for digital preservation with a notable example being a joint project between The Andy Warhol Museum, STUDIO for Creative Inquiry, and the Carnegie Mellon University (CMU) Computer Club. The aim of the project was to access forty floppy disks that potentially contained never-seen works by artist Andy Warhol, created using a 1985 Commodore Amiga 1000 as per a commission by Commodore International to show off the graphic power of their computer. In March of 2013, the Computer Club met with museum staff and brought equipment including two different Amiga computers, drives, and floppy disks owned by the Club, and of course the KryoFlux, enabling images to be made of all but four of the disks. These disk images allowed the Computer Club to access their contents as virtual disks via UAE, an Amiga emulator, to load the images and familiarise themselves with the software Warhol used and reverse engineer some of the file formats to more widely used variants.[22] Currently the Warhol Museum is working with IonTank to "create an Amiga-based interactive that visitors can actually operate to navigate a selection of 10 origin files created by Warhol and his assistants as they learned to use the Amiga 1000."[23]

This example shows the wide-ranging potential of tools such as KryoFlux beyond its primary purpose. It also demonstrates that there is not an all-encompassing solution to digital preservation. Older hardware was still required to read the floppy disks and the use of an emulator combined with knowledge of the Amiga system were all essential components of the preservation process. This also highlights how the different methods of preservation can

[22]"Warhol Amiga Report," The Frank-Ratchye STUDIO for Creative Inquiry, (March 2014): 1-6, http://studioforcreativeinquiry.org/public/warhol_amiga_report_v10.pdf.
[23] Jeffery Inscho, "Hacking Vintage Technology to Simulate Time Travel", The Andy Warhol Museum, Accessed October 26, 2019, https://www.warhol.org/hacking-vintage-technology-to-simulate-time-travel/.

complement one another and that no preservation project has to be limited to utilizing only one. KryoFlux helped to simplify this process and similar projects can be undertaken without fully intact computing legacy environments (which the Museum and Computer Club had) thanks to the bridging of modern and legacy environments afforded by KryoFlux.

KryoFlux is a specialized and undeniably useful piece of hardware but is not a preservation program in and of itself. Security and authenticity (of the data) are essential to ensure confidence that preserved files or objects have not potentially been damaged or tampered with in some way. Because digital records are designed to be manipulated via environments there can be obvious cause for concern when interacting with legacy media, hardware, and software. Older magnetic media is susceptible to corruption and loss of data when accessed and content may be altered during the process. However, it is also impractical to rely solely on legacy systems to analyse recovered data due to their fragility. So how can archivists safely recover and preserve data while also limiting unnecessary interactions without putting legacy environments at risk? The answer is through digital forensics: hardware specifically designed to gather, preserve, and analyse digital information or evidence utilised by law enforcement and investigations.[24]

Digital forensics equipment and software is available for purchase and personal use by institutions and individuals to ensure safe and authentic analysis and preservation of digital content. Even if a user is only looking at the contents on a drive, changes to the contents may still be made automatically by the system simply by plugging it into another computer. Utilizing a basic piece of digital forensics hardware and/or software such as a write-blocker ensures that

[24] Adam Scott Wandt, "Digital Forensics: Hardware," https://www.youtube.com/watch?v=-qPVtaxJWv4&feature=youtu.be.

the drive is truly read-only keeping that data safe and preserving its authenticity.[25] As discussed in the previous chapter recordkeeping systems are meant to ensure accountability, and one of the defining traits of digital records is their ability to be altered. So as the volume of digital records increases it is all the more important to ensure there are reliable tools available to handle records safely and efficiently. Digital forensic tools help to ensure that the media and data within an archive's holdings has not been tampered with.[26] This helps ensure that subsequent versions created through format migration or other means are derived from the same authentic original.

Forensic workstations are sophisticated systems that include multiple hardware and software components dedicated to digital forensics. One could even describe such workstations as digital forensic environments. The Forensic Recovery of Evidence Device (or FRED) can examine various forms of storage media utilizing write-blockers, create disk images, recover deleted content, and can perform automated actions to identify specific information, such as social security numbers, and flag them for removal or restriction.[27] Digital forensic environments such as a FRED enable archivists to engage with legacy environments in a sophisticated and secure manner. The ability to preserve authentic copies of data from legacy media on a bit for bit level[28] rather than immediate format migration means that, while the original legacy environments may not be the primary means of interacting with the data they are still a part of its provenance and can be identified from file format extensions and associated metadata. There is more to access than a running drive, however. The definition of a digital environment in this thesis is derived from Dappert et al.: "the hardware and software required to properly interpret the bit sequences of a digital record and render it for human viewing and use, allowing for

[25] Wandt, "Digital Forensics: Hardware."
[26] Ibid.
[27] Natalie Vielfaure, "Have you Heard About FRED?," *Research Services Digital Strategies Blog*, Posted November 15, 2018, https://libguides.lib.umanitoba.ca/researchservices/rsdsblog/Have-you-Heard-About-FRED.
[28] Vielfaure, "Have you Heard About FRED?"

interaction and potentially modification."[29] By utilizing this data it is possible identify the software and hardware dependencies of digital record formats necessary to render and manipulate them.

Identifying the software associated with older file formats can still be tricky since the standardisation we enjoy today was not always the norm; .doc, for example, does not mean a file is necessarily Microsoft Office Word compatible as this extension has been used differently in the past. Fortunately, institutions are developing databases to aid in identification of formats, such as the online information system PRONOM run by the National Archives of the United Kingdom,[30] which also provides a free identification tool, DROID, that can be used offline and independently of PRONOM.[31] It is clear that digital preservation, just like the preservation of non-digital records and artifacts, requires access to the right tools to get the job done. Despite this Levi reached a somewhat discouraging conclusion: "the Archives of Ontario should not become a museum of obsolete technology; that said, some embracing of the ancient machine will be required for at least the next decade – by which time the storage devices we now think of as common might also be heading towards extinction."[32] If the Archives of Ontario example is anything to go by, then archives may well have to shift some of their focus and policies towards acquiring and preserving older digital environments (i.e. hardware and software) in order to access their own content, or they really will have to seek out other parties in the hope that they will be willing to aid them.

An example would be the efforts of Library and Archives Canada (LAC) to improve upon their current digital preservation methods. Recently LAC has begun working with

[29] Dappert, et al., " Describing and Preserving Digital Object Environments," 110-11.
[30] "PRONOM," The National Archives, Accessed February 11, 2020, http://www.nationalarchives.gov.uk/PRONOM/Default.aspx#.
[31] "DROID Signature Files (Archive)," The National Archives, Accessed February 11, 2020, https://www.nationalarchives.gov.uk/aboutapps/pronom/droid-signature-files.htm.
[32] Levi, "Five Hundred 5.25-Inch Discs," 245-6.

specialists from Preservica, a developer of digital preservation systems,[33] to create proprietary software that can better manage the ingest of large volumes of digital materials.[34] Governments understandably generate a lot of records and as LAC is the official archive for many of them archivists have been working hard to develop tools to help manage archiving them. One method is the creation of Journey Maps to track how a record moves throughout systems and departments and how people interact with them to better understand their functions.[35] LAC has also been in close communication with government departments to better define the importance of front-end involvement in record-keeping to prepare records before they enter into archivists' care.[36] Working with records creators and specialists, when possible, can help in the long run to ease the transition of records into archival custody and develop policies geared towards the long-term preservation of records. Such focus and policies must also account for rapid obsolescence which will require archivists to understand that current environments and media will, too, fall under Levi's category of "the ancient machine", as will their successors, likely within their lifetime. This means that the estimate of at least a decade is too short a time frame. Archivists are still dealing with backlog of digital records on media that was first commercially available in the 1960s (floppy disks) so it is not unreasonable to assume that current environments and media will enter into this backlog as well as the next generation of legacy environments and media. The Archives of Ontario is only one example, but it shows just how important obsolete environments are for any archive or institution that retains any obsolete media in their holdings.

From the Hobbyist Movement to Retrocomputing – The Computer is Personal

[33] "About Us," Preservica, Accessed February 11, 2020, https://preservica.com/about.
[34] Sylvain Belanger, "LAC Digital Optimization of Canada's Collections," Presentation, University of Manitoba, December 2, 2019.
[35] Belanger, "LAC Digital Optimization of Canada's Collections."
[36] Belanger, "LAC Digital Optimization of Canada's Collections."

Access is the primary goal of digital preservation: access to content and to context for current and future users. However, this can be difficult when it comes to handling original hardware and software, as the example of the Archives of Ontario demonstrates. If there is no digital environment an obsolete record may be inaccessible. Su-Shing Chen neatly summarised the practical concerns of digital preservation: "On the one hand, we want to maintain digital information intact as it was created; on the other, we want to access this information dynamically and with the most advanced tools."[37] There is, moreover, more to an artifact than its physical presence; there is the history tied to it. Focusing upon the migration of individual data objects as an ongoing method of preservation may help to lessen the eventual obsolescence of current storage media and records, but it does not adequately convey the importance of environments. This means users and researchers will have little comprehension of the record creation process. Levi's statement that the archive should not become a museum of obsolete technology is also problematic as it suggests that the digital record and environment are wholly independent of one another and that the latter has no place within the archive. This simply is not true because, as discussed in the previous chapter, computers have been important cultural touchstones for decades as symbols of national or organizational technological prowess and points of contention about the future of the workplace. It would of course be unreasonable to expect an archive to accommodate every possible iteration of digital environments that apply to the whole of their digital records. However, this does not mean environments do not have a place within archives in one form or another.

As stated in the previous chapter computers, like anything else, are not created in a vacuum but to fulfill a purpose. Even if the purpose remains the same the form and function of a tool or technology can differ depending on who develops it and where. Patricia Galloway stated

[37] Su-Shing Chen, "The Paradox of Digital Preservation," *Computer* 34 no. 3 (March 2001): 25.

that the history of computing has often overlooked personal experiences and historians "seem to have concentrated on the history of technological development rather than the details of the interaction of people and technology in daily work practices."[38] The adoption and impact of computers, particularly personal computers (PCs), was not a uniform one; they were as much subject to social and political forces as the people who developed and used them.

One of the major driving forces behind widespread PC adoption was the hobbyist movement which originated with radios and early electronics, supported by dedicated magazines as early as 1908 with *Radio-Craft*, and *Popular Electronics*, which launched in 1948, and which grew to encompass early 8-bit microprocessors that first came on the market in the 1970s.[39] To hobbyists the PC, like the radio, was more than a new luxury to place on a table. There was a desire to understand how it worked and to tinker with it, and alter and improve it. The hobbyist movement was not limited to North America and reflected the circumstances that lead to its manifestation in different parts of the world in both the people and the technology they were driven to understand. This gives many different legacy computers unique national identities when viewing them through a geo-political lens.

During the Cold War the CoCom embargo limited the sales of computers from the West and forced manufacturers in the Eastern Bloc to develop unique systems, even reverse engineer and clone Western ones to meet their computing needs.[40] Limitations in manufacturing capabilities, coupled with the embargo, meant that the adoption of PCs within the Soviet Union and Eastern Bloc hobbyist movements occurred nearly a decade after the North American

[38] Patricia Galloway, "Personal Computers, Microhistory, and Shared Authority: Documenting the Inventor–Early Adopter Dialectic," *IEEE Annals of the History of Computing* 33 no. 2 (2011): 60.
[39] Zbigniew Stachniak, "Red Clones: The Soviet Computer Hobby Movement of the 1980s," *IEEE Annals of the History of Computing* 37 no. 1 (2015): 12-13.
[40] Stefano Bodrato, Fabrizio Caruso, Giovanni A. Cignoni, "Discovering Eastern European PCs by Hacking Them. Today," arXiv (2019): 3, https://arxiv.org/abs/1901.06922.

movement.[41] This was one of the unique traits of the Eastern Bloc hobbyist movement that differentiated it from the North American movement; the need to improvise as a means to make the hobby possible was not purely an indulgence. In 1978, utilizing the Soviet produced KR580IK80 microprocessor, and technical literature about the Intel 8080 CPU which it was reverse engineered from, three employees of the Moscow Institute of Electronic Engineering (MIEE) developed the Micro-80 PC. They then published the schematics and operating information in the long running magazine *Radio* in 1981, simply for the purposes of drumming up interest in microcomputing. Instead numerous readers decided to actually build their own Micro-80s even though it meant utilizing the black market to obtain the CPU and other necessary components.[42] From here PC development and hobbyist movements in the Soviet Union and neighbouring countries followed roughly the same pattern. Numerous Eastern Bloc countries developed their own clones of American and British computers, often purposely designed to be compatible with the originals, and targeted at the education and industrial sectors before rebranding them as DIY kits or publishing the schematics in magazines which would, with varying degrees of success, see hobbyist groups form around them.[43]

The hobbyist movements, in particular in the Eastern Bloc, reveal just how personal the PC was for many early adopters, and the impact user agency can have on computer development. Individual tinkering made hardware personal, not just technical, and exemplified the fact that the PC was not adopted solely due to the fact that they were the next logical step in the evolution of computing. Early adopters had a desire to understand how they worked, the same as radios and other early electronics. In the case of the Eastern Bloc this sometimes meant starting from scratch and even utilizing illegal avenues to purchase essential components as a work-around to

[41] Stachniak, "Red Clones," 14-15.
[42] Ibid, 15.
[43] Bodrato, et al. "Discovering Eastern European PCs," 9-10.

political roadblocks.[44] Personal investment in computing is still alive today and has seen new tools developed to keep tinkering alive and well, not only on legacy systems but modern ones as well. These systems may now be obsolete, but interest has not died and is now maintained in a variety of ways due to the continued enthusiasm of original hobbyists and the modern retrocomputing community. Today community-produced emulators and cross-compilers and cross-assemblers allow these enthusiasts to continue to tinker with code on modern systems which is only possible due to the intimate knowledge they possess and time devoted to preserving legacy Eastern Bloc PCs.[45] And while these compilers are not full system emulations they allow users to experience select functions of older systems, demonstrating how legacy computing and (limited) emulation can be utilized in tandem as a means of preserving digital environments. These compilers and assemblers are also useful for authenticating and better understanding clones since they can be utilized to identify portions of firmware code derived from Western CPU architecture and modified to be compatible with Eastern Hobbyist systems.[46]

Such tools for preservation, identification, and continuing interest in coding and tinkering are only possible because the original hobbyists and more recent retrocomputing enthusiasts have taken the time to hack, disassemble, and tinker with older systems. What is also being preserved along with these legacy environments is the context in which they were made; unique digital cultures born out of Cold War era politics and the universal desire to tinker with and understand new technologies. What is also being preserved is the "tacit knowledge" that is inherent only through physical interaction with the systems themselves.[47] Hardware and software are cultural

[44] Bodrato, et al., 8.
[45] Ibid, 11-12.
[46] Ibid, 14.
[47] Patricia Galloway, "Retrocomputing, Archival Research, and Digital Heritage: A Computer Museum and iSchool Collaboration," *Library Trends* 59 no. 4 (Spring 2011): 625.

artifacts. This means that something like a floppy drive, which on its own may seem mundane, is almost certainly tied to a larger cultural or subcultural movement.

While obsolescence means older computers are retired and removed to be replaced it does not mean they are totally abandoned by society or deemed to have lost all cultural value. They are still worthy of ownership and use in the eyes of many. Such users can range from institutions such as archives or dedicated museums, to individuals[48] or small organisations working with limited resources. The continued utilization of older or obsolete computer systems, including interest in the use and preservation of older computer systems, can be labelled broadly as legacy computing. Legacy computing is usually the result of the usual hurdles associated with upgrading including cost, lack of desire or need to change the current system, or the sheer scale – particularly in the case of large businesses – involved in upgrading technological infrastructure.

On an individual level an interest in preserving and tinkering with legacy and obsolete systems is often identified as retrocomputing. Patricia Galloway identifies three defining criteria of retrocomputing: "(a) amateurism, in that the time spent on legacy computing is not generally compensated … (b) technological skill acquired through education, professional experience, or personal study; and (c) the persistent interest that springs from a sincere identity with and interest in the field."[49] Retrocomputing is often hobby-based and fuelled by a passion for understanding and preserving older hardware and software for personal interest and even public display. It is an unprecedented area of preservation because of how (relatively) young digital computing is. As Galloway notes: "Almost the whole of [digital] computer history is a temporal target from which

[48] Some of the most successful authors utilize obsolete environments in their work. George R. R. Martin, of *A Song of Ice and Fire* fame for example, utilizes WordStar 4.0 on a machine running DOS to avoid spell check and autocorrect interfering with his work. "George R.R. Martin Writes Everything In WordStar 4.0 On A DOS Machine," TechCrunch, Accessed February 13, 2020, https://techcrunch.com/2014/05/14/george-r-r-martin-writes-everything-in-wordstar-4-0-on-a-dos-machine/.

[49] Galloway, "Retrocomputing," 625.

living expertise is still available – explicit documented knowledge, yes, but also the tacit knowledge embedded in the design and use of both hardware and software."[50]

Retrocomputing, despite its hobby status, is a major force for preservation in its own right and is built upon the same passion to tinker and understand as the hobbyist movement of the 1970s in North America and 1980s in the Eastern Bloc. The primary difference is that the objects of their focus are obsolete, but this is not indicative of expertise:

> And to the extent that retrocomputing participants are "hobbyists," in the sense of not being paid for their work, they are hardly unskilled amateurs. Rather, their practice often demonstrates deep sophistication. In other words, many of them are "hobbyists" only in the same sense as many of the contributors to open source software, which today underlies much of the world's computing infrastructure.[51]

Retrocomputing is about preserving computer culture but, as Takhteyev and DuPont point out, contributors are well versed in computer nomenclature and technical know-how, which often leads to creative work-arounds to issues of aging and deteriorating systems. Retrocomputing is often

> transformative because simple preservation is not the participants' only (or even main) goal. Instead, retrocomputing enthusiasts engage with old computing for a variety of reasons, including a simple desire to have fun with old technology … they often seek to adapt systems to their contemporary needs – for example, finding ways to connect old computers to newer ones via modern networking.[52]

This willingness, desire even, to adapt, aids in preserving older environments by allowing the obsolete to function within a modern space while still retaining its original functionality and distinguishing characteristics.

Collaboration

While locating original hardware and individuals with the requisite knowledge to competently repair (or even use) older environments can prove difficult, archivists are fortunate

[50] Ibid, 625.
[51] Yuri Takhteyev and Quinn DuPont "Retrocomputing As Preservation and Remix," *Library Hi Tech* 31 no. 2 (2013): 357.
[52] Takhteyev and DuPont, "Retrocomputing," 358.

when there are users of original systems who can share their technical and cultural expertise and experiences. This has led to the development of tools such as the ones utilised by Western hobbyists and their Eastern Bloc counterparts. Collaboration is a perfectly valid solution to issues concerning preservation of older environments since it not only allows access to equipment and expertise, but also serves as a testimony of their value.

In 2007 students from the University of Texas School of Information started volunteering at the Goodwill Computer Museum, drawn in by mutual interest in the legacy machines and a chance to apply their archival, library, and information science education.[53] Over time this relationship enabled the School of Information and Goodwill Computer Museum to pool resources and undertake joint preservation projects that would have otherwise been impossible. The Goodwill Computer Museum had numerous artifacts and forms of documentation, both digital and printed, so when members wished to perform in depth cataloguing it was an opportunity for School of Information students to familiarise themselves with industry nomenclature and research materials which they were then able to translate into a preservation program.[54]

This collaboration benefited both parties as they are able to exchange experience and services and address issues of preserving computer culture. It can also serve as a model for other groups and institutions to use as a frame of reference for developing or expanding their own programs. The Archives of Ontario had a collection of legacy media but initially lacked the hardware to access it. The School of Information encountered a similar issue in 2009 but it grew into another collaborative project with the Goodwill Computer Museum. They custom-built a machine dubbed Frankenstein I (later substituted with the more capable Frankenstein II in 2011)

[53] Galloway, "Retrocomputing," 628.
[54] Ibid, 631.

using legacy hardware and drivers programmed by Museum volunteers with input from students and the university archivist for the purposes of capturing legacy files, software, and relevant information from a variety of storage media via disk imaging to create authentic digital objects.[55]

The sharing of tools and expertise is important because it enables archivists to acquire the knowledge and develop the skills needed to better understand and interact with older digital environments. Preservation projects, such as Frankenstein I and Frankenstein II, are important because they have the long-term benefit of preparing future archivists for the eventuality that even the most current systems will attain legacy status. Of course, the preservation of hardware and software is not limited to decades old equipment. In fact, it is of paramount concern right now. Today's laptops and applications are tomorrow's obsolete tech and will present archivists with the same problems as floppy disks and magnetic tape. Modern environments will also present more immediate concerns; specifically, developing the familiarity and technical know-how to operate and understand them. Without the means to pass on the necessary knowledge, both documented and tacit, the issues of obsolescence will only be compounded by current environments as they age.

Methods of Preservation 3: Emulation

The fragility of legacy systems means that preserving them indefinitely is not practical. This is why many institutions choose to preserve only the hardware and software required to transfer data to newer storage media before moving on to format migration. Since software can be copied to newer media one solution is to move away from dependence upon legacy hardware through emulation. Simply put, "emulation recreates a computer environment (target) on top of another computer environment (host)."[56] A more technical description would be that the purpose is to emulate older software applications, hardware, operating systems, drivers, and

[55] Ibid, 632.

other components to allow aspects of the older environment to be experienced on current systems.[57] The most noticeable difference, from a user's perspective, is that emulation cannot reproduce the full tactile experience of utilizing a legacy system, as retro and legacy computing can. Files and application software are loaded with the click of a mouse rather than physically inserting a floppy disk into a drive, all while the user stares into a modern high definition flatscreen monitor. Such disparities in experience between original and emulated environments are inevitable due to the absence of original hardware. This move away from dependency does strike a parallel with format migration. However, the purpose of emulation is to preserve and make accessible as much original functionality as possible, for both environments and, in turn, records. As stated above, preserving actual hardware indefinitely is impractical. Emulation enables developers and users to work around this by replicating physical hardware through programming. This makes it possible to potentially maintain emulated environments indefinitely since they can rely upon updated coding rather than finite physical resources to remain accessible.

Emulation is resource intensive and requires in-depth technical and tacit knowledge in order to make legacy environments accessible on modern computer systems. At the same time emulation is essential for accessing and rendering complex digital objects such as video games, computer generated artwork, and datasets.[58] As previously stated, a digital record is a combination of processes that serve to interpret and display data in some meaningful way. This description can be expanded to include more complex digital objects such as applications software including video games. As complexity increases so too does the task of preservation to

---

[56] Jeffery van der Hoeven, "The Need for Emulation Services," *De Gruyter* 35 No.4 (2012): 235.
[57] Jeffery van der Hoeven, Bram Loham and Remco Verdegem, "Emulation for Digital Preservation in Practice: The Results," *The International Journal of Digital Curation* 2 No. 2 (2007): 126.
[58] Dianne Dietrich, Julia Kim, Morgan McKeehan, and Alison Rhonemus, "How to Party Like it's 1999: Emulation for Everyone," *code[4]lib Journal* 32 (2016). https://journal.code4lib.org/articles/11386.

the point that format migration is no longer an option. The National Library of the Netherlands (KB) has a collection of niche educational and scientific multimedia applications that rely upon specific systems in order to be accessed which makes conversion impossible.[59] Since the combination of processes is environment-dependent this means that emulation is the only practical means of maintaining accessibility to complex digital objects beyond the life of the original physical environment. While emulation cannot preserve the tactile experience of utilizing legacy environments like retro and legacy computing it does afford users the opportunity to experience some of the functionality of older systems and records.

Authenticity and Accessibility

Since emulation is a difficult undertaking it is important for a given archive to figure out exactly what it is working with and what it needs. Digital environments have both technical and cultural significance but the blunt truth is that most archives simply do not have the time and resources to emulate and preserve digital environments. One of the most important sources of information are the records themselves since they are a major determinant of whether emulation is warranted. By identifying potential candidate materials and their dependencies it is possible to identify an environment that can integrate into an archives' current digital preservation workflow. Simply put, "The more you can determine about that environment, the better you can figure out what emulation strategy will work best for your materials."[60] Presentation is one of the aims of emulation by ensuring the correct combination of processes produce a performance that displays the record within emulated environment correctly.

Emory University's Manuscript, Archives, and Rare Book Library (MARBL) determined emulation was a more "holistic" means of preserving and presenting some of the works of author

[59] Hoeven, "The Need for Emulation Services," 236.
[60] Dietrich, et al., "How to Party Like it's 1999."

Salman Rushdie because "If one emulates an obsolete operating system, the concerns about loss, authenticity, and error that plague migration largely disappear."[61] Since MARBL had Rushdie's personal Mac Performa this enabled them to create disk images to aid in modifying the emulation to produce a more authentic performance of Rushdie's personal computer down to directory structure, user preferences, and more. While this allows users to modify and remove files and directories and play the games available on the original system, such changes are not permanent and the whole environment defaults back to the disk image upon restarting.[62] This raises two questions: what manner of authenticity should be applied to emulation; and when does it potentially clash with practicality? Digital Forensic Analyst Dianne Dietrich states that "Forensic authenticity and archival authenticity refer to different ways of ensuring that an object is what it claims to be, and has not been falsified or corrupted."[63] In this regard the MARBL project is arguably successful since users can access a fairly faithful recreation of Rushdie's computer and records, along with their functionality. This is also possible without concern for users making permanent changes in the course of their interaction. From a practical perspective developing single use emulation solely to replicate an individual's system places severe limitations upon a powerful tool for preservation. Jordan Roy notes that "The MARBL project was made more feasible by the circumstance of only having to set up one instance of emulation on one computer … There would be no worry about needing to adapt the emulation to various browsers and OS environments."[64] Emulation is most practical when it can be applied to records

[61] Laura Carroll, Erika Farr, Peter Hornsby, and Ben Ranker, "A Comprehensive Approach to Born-Digital Archives," *Archivaria* 72 (2011): 79.
[62] Carroll, et al., "A Comprehensive Approach to Born-Digital Archives," 84-5.
[63] Erin Engle, "Authenticity Amidst Change: The Preservation and Access Framework for Digital Art Objects," *The Signal,* Library of Congress, Posted December 21, 2015, https://blogs.loc.gov/thesignal/2015/12/authenticity-amidst-change-the-preservation-and-access-framework-for-digital-art-objects/.
[64] Jordan Roy, "Preserving Interactivity: Towards Next Generation Digital Preservation Philosophy and Systems," Master's Thesis (University of Manitoba/University of Winnipeg, 2019), 84. MSpace

across multiple fonds and workstations which makes single use projects meant to exist in such isolation less than ideal.

A better example of balance between authenticity and practicality would be the ongoing efforts by the Internet Archive to preserve legacy environments and applications software. Formed in 1996 with the focus of archiving the Internet the Internet Archive has expanded its preservation efforts to include, beyond the archived 330 billion webpages, over 200,000 software programs amongst a variety of other content.[65] Whereas MARBL's emulation was limited to a single workstation and built around a highly personalised instance, the Internet Archive has chosen to provide a more widely accessible stock experience. On February 11, 2016 and April 16, 2017 the Internet Archive announced the addition of browser-based emulations of Microsoft's 1992 release of Windows 3.1[66] and the 1991 release of MacOS System 7.0.1 and respective software applications.[67] This browser-based form of accessibility allows for users to interact with their collection without any technical expertise or having to endure complicated installation procedures. All of the technical heavy lifting has been done by the Internet Archive and contributors to configure and port the necessary emulators and applications software into JavaScript so that they can function within a user's web browser.[68] This is by no means the sum of the Internet Archives efforts to leverage emulation in such a manner. The Windows 3.X and MacOS emulations are additions to an extensive number of software libraries available for users including Atari, Commodore 64, Apple II, MS-DOS, and more.[69]

---

[65] "About the Internet Archive," Internet Archive, last modified July 12, 2019, https://archive.org/about/.
[66] Jason Scott, "Internet Archive Does Windows: Hundreds of Windows 3.1 Programs Join the Collection," Internet Archive Blogs, last modified April 04, 2018, https://blog.archive.org/2016/02/11/internet-archive-does-windows-hundreds-of-windows-3-1-programs-join-the-collection/
[67] Jason Scott, "Early Macintosh Emulation Comes to the Archive," Internet Archive Blogs, last modified April 04, 2018, https://blog.archive.org/2017/04/16/early-macintosh-emulation-comes-to-the-archive/
[68] Scott, "Internet Archive Does Windows."
[69] "Software Library," The Internet Archives, Accessed May 11, 2020, https://archive.org/details/softwarelibrary.

The Internet Archive is an example of how emulation can be implemented in a manner that makes it widely accessible, to a degree. While users can experience authentic performances of legacy operating systems and applications software from their browser on computer anywhere it cannot be readily expanded to include digital objects from archival holdings beyond the Internet Archive's servers. To do so requires users to do some technical heavy lifting on their own. Where it is possible in many cases to download the necessary files to set up a local installation, users will have to perform the configurations on their own. From a general practical perspective what the Internet Archive is accomplishing is beyond the capabilities of most archives who cannot divert most of their limited resources, which are also dedicated to other materials, to maintain large emulation projects. Roy rightly notes that "Unless the amount of digital content is small, the number of digital objects being adapted for emulation is low, have a dedicated digital preservation and access staff, or, for whatever reason, [the archives] maintain a large, highly diverse, and specialized workforce, this approach is not generally feasible."[70] For archives wishing to implement and integrate emulation as part of their workflow the Internet Archives approach also presents the problem of not allowing access beyond the emulation. Users may load an instance of WordPerfect but cannot import, save, or export files as the emulated applications software exists on its own separate from any environment. The Internet Archive does "allow convenient access to a large collection of historic computer games but lack[s] portability and generality … The emulation instances remain stateless and accessing user data and customization is not supported."[71] Fortunately there have been efforts to make emulation

[70] Roy, "Preserving Interactivity," 88-9.
[71] Thomas Liebetraut, Klaus Rechert, Isgandar Valizada, Konrad Meier and Dirk von Suchodoletz, "Emulation-as-a-Service – The Past in the Cloud," *2014 IEEE 7th International Conference on Cloud Computing,* Anchorage, AK, (2014): 906. 10.1109/CLOUD.2014.124.

adoption and implementation easier for individuals and archives wishing to make their legacy materials accessible through emulation.

Automation and Virtual Machines

Large collections of legacy digital media will almost certainly contain some form of legacy applications software or executables that are integral to accessing encoded data which cannot be migrated without the threat of data loss.[72] After emulating an environment there is still the matter of installing the necessary applications software or executables in some manner in order to access the data. One could simply have all the necessary applications software and executables installed by default every time the emulation runs but this could make it cumbersome if only one is needed. There is also the option to install them as needed within the emulated environment but this can be complicated and users may simply not know how to accomplish this on older and unfamiliar systems.[73] Alternatively each separate executable and application could be handled as an independent emulation, like at the Internet Archive, but this makes more work and also dilutes the functionality and experience of using a complete emulated environment. This solution would be both impractical and cumbersome for developers, archivists, and users. Fortunately, there has been much work over the years to make emulation more adaptable and streamline the installation process.

Because original hardware cannot be relied upon for long-term preservation, yet is still an essential part of how a digital environment functions it must be replicated in some manner. This is accomplished by developing a virtual machine (VM) which is "a computer that has no separate physical existence, but is part of the behavior of a physical computer, called the virtual machine's host computer. Virtual machines mimic the instruction set and hardware configuration

[72] Kam Woods and Geoffrey Brown, "Assisted Emulation for Legacy Executables," *The International Journal of Digital Curation* 1 No. 5 (2010): 161.
[73] Woods and Brown, "Assisted Emulation," 161-2.

of some physical machine."[74] More simply, a virtual machine is essentially an artificial second computing environment that can run within a physical one, or host. In order to create a VM you need information about the environment you are trying to replicate. This information is often obtained from ISO files or images, also know as disk images, which are exact, bit-for-bit copies of digital objects, software applications, operating systems, or any other digital content. These can be obtained through disk imaging software.[75] ISO images are convenient because they contain executable files which may contain instructions such as initiating the installation of applications software. This is a boon to the automation of emulation and VM development and implementation since they can be stored – along with any essential accompanying data – in unique directories and called upon as needed.[76] Utilizing ISO images also provides stability and security since "The ability to mount disk images as specific virtual devices, and beginning each installation in a "clean" virtual environment are critical components for successful access to many legacy executables."[77] By preserving a stable "snapshot" of an operating system and applications software there is always a functional point of reference to return to in the event of errors or some form of corruption. Any glitches that may occur through bit-rot or are inherent to the legacy environment do not mean potential doom for an emulated environment, as they might be for an original legacy system. You can utilize the eternal go-to solution of "turning it off and on again", after a fashion, by resetting the VM then re-loading the software.

From the point of view of an end-user all that is required is to identify what sort of environment is necessary to reproduce a record, and then select the appropriate operating system

[74] David S. H. Rosenthal, "Emulation & Virtualization as Preservation Strategies," LOCKSS Program, Stanford University Libraries (2015): 2. https://digital.library.unt.edu/ark:/67531/metadc799755/m2/1/high_res_d/rosenthal-emulation-2015.pdf.
[75] Tim Fisher, "What is an ISO File?," *Lifewire*, Last Updated April 13, 2020, https://www.lifewire.com/iso-file-2625923.
[76] Woods and Brown, "Assisted Emulation," 163.
[77] Ibid, 165.

and software to load (mount) within a virtual machine. While this sounds convenient it is still a significant technical undertaking to make a VM function within an emulated environment in order to reproduce an authentic performance of a legacy environment within a modern one. Rosenthal explains that when running an emulation "None of the instructions executed by the VM are directly executed by the host computer's CPU; all are translated by host computer software from the VM's instruction set to the host computer's instruction set by host software before being executed."[78] This translation process is what makes an emulator possible. It allows for a modern environment to function as an older one – software-wise – via an advanced form of mimicry. Therein lies much of the complexity behind emulation; it not only requires rebuilding much of the original environment but also developing the necessary compatibility with newer ones to ensure translation – and therefore mimicry – are possible across multiple systems.[79]

Emulation-as-a-Service (EaaS)

I have stated earlier in this thesis that no method of preservation discussed in this paper is meant to be an all-purpose solution to preserving digital environments, rather, it is a matter of individual archives determining what combination is most practical for them. This will be dictated not only by the resources and expertise available to archivists but also the nature of the records in their care, and the mandate of their institution. If emulation is to gain more widespread adoption as a method of preservation, the means to produce and deliver quality emulated environments must be flexible and adaptable enough to meet the demands of different archives. This is where Emulation-as-a-Service (EaaS) comes in.

[78] Rosenthal, "Emulation & Virtualization as Preservation Strategies," 3.
[79] To be clear, it is the emulator that is *compatible* with one or more modern systems. Records in their original format which cannot be opened in other programs or migrated to accessible formats are therefore *dependent* upon the original environment in which they were created in order to be rendered. Emulation is a sophisticated workaround to the issue of dependence by making an incarnation of a legacy environment that can be run on a modern one.

Baden-Württemberg Functional Long-Term Archiving and Access (bwFLA), developed by the University of Frieberg, is one such example. Similar to the Internet Archive, users are able to interact with the emulation though their web browser on their personal computer after selecting a link and wait to be redirected after the appropriate virtual machine. Processing power and other resources are allocated server-side.[80] The goal is to centralise the resources and expertise required for emulation and dedicate them to handling the most difficult technological challenges in order to develop a sustainable model by which emulation can be provided quickly and efficiently to a large number of users. Limiting the diversity of hardware allocated to run the emulations server-side makes general maintenance more convenient but there is still the matter of making it accessible to clients

> As computer systems change in many aspects over the time, the API [Application Programming Interface] needs to be able to translate and to emulate the several forms of in- and outputs. This could require the addition of, for example, an overlay keyboard or game console controller, or the provision of a virtual mouse on touch screen devices.[81]

People are used to accessing materials from a variety of devices such as desktops, laptops, tablets, and phones. This means that ongoing development is still essential to both adapt to new technologies and work to include a wider variety of established ones. While accessing a full desktop environment on small screen devices may not necessarily be practical it could potentially pave the way for the possible preservation and reproduction of environments originally developed for mobile devices.

In 2015 bwFLA developed the option to download and run preconfigured environments locally through a graphical user interface (GUI) of dropdown menus and included additional

[80] Rosenthal, "Emulation & Virtualization as Preservation Strategies," 6.
[81] Dirk von Suchodoletz, Klaus Rechert and Isgandar Valizada, "Towards Emulation-as-a-Service: Cloud Services for Versatile Digital Object Access," *The International Journal of Digital Curation* 8 no. 1 (2013): 138.

supports to develop and monitor workflows for ingest and customization.[82] This means that institutions have greater personal control over how they can integrate emulation. Bottlenecking due to bandwidth is also less of a potential issue for in-house use. However, this means foregoing the benefits of the public cloud such as requiring the outright purchasing and ongoing maintenance of their own servers and installations, both of which will experience periods of underutilization.[83] Regardless this still makes emulation more attainable for archivists that do not have the technical knowledge or experience to develop emulation and institutions that would otherwise avoid such projects due to lack of resources.

As a concluding point it is important to remember that today's modern environments are tomorrow's obsolete legacy environments, and they are only growing more complex and sophisticated. Emulation has the potential to become a mainstay in digital archiving, one that can better preserve the functionality and context within the record-environment relationship. Dietrich et al. nicely summarise that

> While archives are just beginning to address backlogs of born-digital content for preservation and access, as more archival collections continue to collect contemporary (born-digital) collections, understanding the problems learned from emulating yesterday's obsolete collections now will lay the groundwork for the future.[84]

Methods of Preservation 4: Documentation and Description

There is the probability that preserving a digital environment, either in physical or emulated form, is neither a possible, practical, nor desirable project for an archive to undertake. Should this be the case, it does not mean that the record-environment relationships within a fonds cannot still be documented through supplementary materials and description, even if the functionalities of the environment are not retrievable in such a manner. This can be

[82] Dietrich, et al., "How to Party Like it's 1999."
[83] Suchodoletz, et al., "Towards Emulation-as-a-Service," 136.
[84] Dietrich, et al., "How to Party Like it's 1999."

accomplished by including relevant information in the form of pre-existing or newly generated reference material. Original manuals, metadata, and other sources of information could be utilized and archival descriptions could be updated to accommodate this. This is not a solution to the problem of preserving digital environments because the fact is that there is no environment being preserved. However, it is still preferable to have some evidence of its existence than nothing at all. Had the Sistine Chapel frescos been lost during restoration, for example, any photographic evidence would be invaluable.[85] Such information is only supplementary and meant to provide users with references, or links to references, regarding aspects of the environment and the potential functionality it would afford records in their original formats.

Why then discuss a method of preservation if it does not actually preserve digital environments? The simple answer is because it provides an archive with at least some means of including, at the very least, the concept of a record-environment relationship and some sort of visual record of what it looked like which could potentially demonstrate some functionality given the right visual media. This is not an alternative to legacy computing and emulation but is well within the capabilities of any individual. Such sources include: archivist-created metadata; original materials (such as user manuals) pertaining to the system; dedicated publications such as magazines and other third-party analysis; and any other related materials. The internet is also a valuable source of information, especially video content such as dedicated YouTube channels where both experts and enthusiasts have made their knowledge widely accessible. Including such sources is also an important part in preserving computer culture.

Archival description will also play an invaluable role since this is often the primary means by which users first interact with records and materials, usually through finding aids. However, the format, adoption, and implementation of the standards that support archival

[85] Though an operating system admittedly differs wildly in form and function from a fresco.

description are not uniform and many do not always accommodate digital (sometimes still referred to as "electronic") records very well, let alone environments. This means that in order to better accommodate environments adaptations will likely be required. Of course archives should not be expected to undertake such accommodation for every file format or operating system that may be relevant to the materials within their archival holdings. This would inevitably grow out of hand. Rather, it would be best to first identify records based on preponderance or to meet the needs of high-value record sets.

Original and Alternative Documentation

In the absence of the environment it falls to archivists to search for sources of information users could reference to get some sense of how the migrated versions of records they are currently examining may have functioned when rendered in their original environment. Two broad categories would be original and alternative documentation. The first emphasises manufacturer produced manuals, guides, and related materials. Alternative documentation includes efforts by individuals and other third parties to describe and preserve information concerning computing environments. It is important to note that efforts to collect such materials should not be limited solely to systems and software that are currently obsolete (such as Apple IIs or Commodore computers). Current environments should also be considered since they will soon become just as obsolete. Therefore, identifying more modern record-environment relationships within one or more archival fonds and collecting relevant material *now* will prove useful for the near future.

Original documentation would primarily exist in the form of records detailing the specifications and operation of specific computing systems and programs. For the sake of practicality archives would have to rely upon a combination of physical and digital copies of

such materials. Physical copies have the benefit of remaining within the control of the archive and can be scanned in if time and resources permit. Developing a collection of such materials could also prove useful for potential future collaborative projects or to cater to enthusiast users. Relying upon physical copies means spending money and time to acquire them and setting aside valuable storage space. There are sites dedicated to selling legacy computing literature, such as oldcomputerbooks.com, which offers a variety of texts and manuals such as Apple user guides form the 1980s for Apple II DOS and Appleworks Calc, File, and similar programs.[86] The Internet Archive is teeming with free accessible scans of similar materials for laptops, gaming consoles, printers, and peripherals such as disk drives.[87] There are also dedicated institutions that have amassed collections of relevant materials, such as the Computer History Museum in Mountain View, California, which possesses large collections of original documentation of legacy computing. Some is also available online in PDF format such as software catalog for Microsoft BASIC from 1980 including commands and features.[88] The existence of such institutions also means there is the possibility of collaboration beyond simply linking to specific materials in archival descriptions. This could include educating users about some of the technologies involved in records creation and their historical significance.

The issue with original documentation is that for the majority of users much of it will be almost impenetrable due to the highly technical language. The larger majority of users will only be interested in the migrated versions of records that are available rather than their environmental provenance. This can make it difficult for many users to engage with such documentation

[86] "Results: Apple," oldcomputerbooks, Accessed April 07, 2020, https://www.oldcomputerbooks.com/advSearchResults.php?action=search&orderBy=relevance&category_id=0&keywordsField=Apple.
[87] "The ARCHIVE.ORG Manual Library," Internet Archive, Accessed April 07, 2020, https://archive.org/details/manuals?&sort=-downloads&page=1.
[88] "Microsoft Software Catalog," Computer History Museum, Accessed April 08, 2020, https://www.computerhistory.org/collections/catalog/102665404.

because it is static material but in reality the record-environment relationship also incorporates interactivity. Even basic user guides to software meant for the average user are still limited by the fact that they are text and images trying to describe an interactive experience. It can be difficult to convey what it was actually like to interact with a digital record in its original format through the necessary environment. The alternative is to look towards different media forms such as video to provide a more vibrant form of documentation. Such alternative documentation can provide users with a sense for original record-environment relationships but in a more dynamic fashion.

Increasingly the Internet is being used as a means for individuals to produce their own instructional content including self-help guides and reviews on almost any topic. Businesses and institutions also utilize the internet as a means to provide users with instructive and informative content based upon their products and services. Some of the most straight-forward examples are uploaded footage from VHS tapes with programs originally directed at then-current users of what is now legacy technology. If a user going through a fonds had an interest in seeing what using WordPerfect was like a quick search on YouTube would provide a list of relevant videos, including a digitized copy of the instructional VHS recording of *Getting To Know: WordPerfect 6.0 for DOS - VHS (1993)*.[89] There also exist uploads of a former PBS series, *Computer Chronicles*, which has an episode covering the same topic.[90] More sophisticated examples would include screencasts where individuals capture their work either through recording their monitor via a camera or directly through software. Since these are often passion projects done by modern day hobbyists archivists may be required to view several videos before finding one that is

[89] "Getting To Know: WordPerfect 6.0 for DOS - VHS (1993)," YouTube video, posted by "member berries", August 10, 2015, https://www.youtube.com/watch?v=jCVGRePLrCA.
[90] "The Computer Chronicles - Secrets of Word Perfect (1992)," YouTube video, posted by "The Computer Chronicles," August 1, 2013, https://www.youtube.com/watch?v=kHel_aqMOb8.

informative enough to meet their needs. This could mean either hour long or multi-video series, or short tutorials such as a fourteen minute MS-DOS tutorial in which the content creator briefly discuses using commands, directories, loading a game, and viewing and copying information from a floppy disk to a hard drive.[91] The obvious concern would be the amount of time required to locate and assess such materials could potentially be prohibitive depending upon the obscurity of the environment or components or applications software. The opposite is also true in the case where there may be too much content to view and appraise.

Archival Description

Whether gathering relevant documentation or actively preserving actual environments, users ultimately have to be made aware of what they can view and why it is significant enough to exist within an archival fonds. This is the role of archival description. The Society of American Archivists defines archival description as "the process of capturing, collating, analyzing, and organizing any information that serves to identify, manage, locate, and interpret the holdings of archival institutions and explain the contexts and records systems from which those holdings were selected."[92] It is a very comprehensive process to make or identify some form of order to apply to records and materials and then describe them in a manner that is consistent and makes sense to potential users of archives. In order to accomplish this a number of reference models and descriptive standards have emerged and continue to develop as guides for archivists. Adoption and implementation are not uniform and differences in the degree of detail and how information is arranged within individual standards means that there is no definitive method of archival description.

[91] "DOS Tutorial Commands for Beginners," YouTube video, posted by "PhilsComputerLab", October 7, 2014, https://www.youtube.com/watch?v=MWiOeV-Fh7U.
[92] Victoria Irons Walch, Compiler, Standards for Archival Description: A Handbook, (The Society of American Archivists: 1994), Introduction, http://www.archivists.org/catalog/stds99/index.html.

Descriptive standards are also not the same as technical standards or specifications, they are open to interpretation and adaptation. Whereas technical specifications "such as those of the Internet Protocol Suite, which enable the flow of data across heterogeneous computer systems and networks, require exact compliance, without which participation in the larger network simply is not possible."[93] Very different from such technical specifications are the descriptive standards used by archivists. These function as guidelines which can, and are, adapted according to the needs of individual institutions. In other words, if an archive chooses to modify or adopt a different descriptive standard from another, this does not stop users from discovering or accessing their records and materials. The purpose of these standards is to provide a framework for ordering information about record sets and materials in a way that makes sense for archivists and users. Implementing these standards does not provide total uniformity however since a number of factors can influence how different institutions approach them including system requirement and decision making at the local level.[94]

Much, if not all, of the information included in archival description such as dates, custodial history, and physical specifications can be classified as metadata. A frequently utilized definition of metadata is "data about data" however this is far from comprehensive. Pat Franks and Nancy Kunde provide a slightly more comprehensive definition by describing metadata as "data describing context, content, and structure of records and their management through time."[95] They argue that metadata is valuable because it "describes, explains, locates, or otherwise makes it easier to retrieve, use or manage an information resource."[96] This also coincides with McKemmish's statement that metadata is present in any materials that describe

[93] Greg Bak, "Trusted by Whom? TDRs, Standards Culture and the Nature of Trust," *Archival Science* 16 (2016): 383.
[94] Bak, "Trusted by whom?," 385.
[95] Pat Franks, Ph.D., CRM, and Nancy Kunde, CRM, CA, "Why METADATA Matters," *The Information Management Journal* Vol. 40 No. 5 (2006): 56.
[96] Franks and Kunde, "Why METADATA Matters," 56.

records and the relationship between them such as registers and finding aids. These definitions are useful but broad. Fortunately, metadata can be broken down into more distinct categories to help identify the different types of metadata and their purpose. These categories include:

- **Administrative:** used in managing and administering information resources;
- **Descriptive:** used to describe or identify information resources;
- **Preservation:** related to the preservation management of information resources;
- **Technical:** related to how a system functions or metadata behave;
- **Use:** related to the level and type of use of information resources.[97]

Rich metadata is invaluable for thorough archival description but developing distinct categories does not make the tasks of creating, collecting, and attributing metadata easy. Institutions also utilise standards to generate and organise metadata, either following in-house or more widely adopted ones, based upon similar factors that influence the implementation of descriptive standards.

Mary S. Woodley notes that user demand for the ability to search for content across different metadata structures "has motivated institutions to convert their legacy content developed for in-house use to standards more readily accessible for public display or sharing or to provide a single interface to search many heterogeneous databases or web resources at the same time."[98] Finding aids and similar descriptive tools are developed by utilizing metadata and in turn are also sources of metadata. This means that archival description is also subject to the benefits and limitations of these standards including the consequences of moving data from one standard to another. Crosswalks are visual representations of the relationship between elements across different metadata schemes and allow search engines to locate similar information across

[97] Anne J. Gilliland-Swetland, "Setting the Stage," in *Introduction to Metadata: Pathways to Digital Information*, 3rd ed., Revised by Murtha Baca (Los Angeles: Getty Publications, 2016), http://www.getty.edu/publications/intrometadata/setting-the-stage/. Indirect quotation, emphasis added.
[98] Mary S. Woodley, "Metadata Matters: Connecting People and Information." in *Introduction to Metadata: Pathways to Digital Information*, 3rd ed., Revised by Murtha Baca (Los Angeles: Getty Publications, 2016), http://www.getty.edu/publications/intrometadata/metadata-matters/.

different databases and aid in converting data to more widely accessible formats.[99] While it is an essential method for migrating data, problems can still occur when undertaking crosswalking. All standards are different – from their internal definitions to what organizational structure they follow – and misalignments can occur such as information being duplicated or obscured, even lost.[100]

While these limitations are not insignificant they do not mean that descriptive standards are irrelevant. The degrees of consistency in implementation between institutions may vary but these standards still provide common points of reference and enable institutions to develop a framework for archival description. The fact is that these standards are not immutable and will very likely be subject to necessary changes or revisions to better support the needs of users by better integrating the search tools available to them.

OAIS and the Information Object

Beyond description there is still the need to collect, organise, and make materials and descriptions, available to users in the long-term. The Reference Model for an Open Archival Information System (OAIS) serves as a high level reference model for digital preservation and provides a common vocabulary for a diverse range of professional groups.[101] It is not as granular as descriptive standards which are formed around managing objects and records at the item level. Rather, it is a reference model which details overarching concepts. The OAIS does not contain instructions describing how such concepts should be implemented. Instead, "from its initial conception, the OAIS was intended to serve as the basis for further development of more specific

[99] Woodley, "Metadata Matters."
[100] Ibid.
[101] Christopher A. Lee, "Open Archival Information System (OAIS) Reference Model," *Encyclopedia of Library and Information Sciences, Third Edition* (2010): 4021.

archives standards."[102] One such standard is the PREMIS Data Dictionary (PREservation of Metadata: Implementation Strategies) which was built upon the OAIS as a foundation for structuring preservation metadata, the "information a repository uses to support the digital preservation process", associated with information objects.[103]

Two of the most significant concepts articulated within the OAIS are the Information Object, mentioned above, and the Functional Model. The Information Object is at the heart of the OAIS. It "is composed of a Data Object that is either physical or digital, and the Representation Information that allows for the full interpretation of the data into meaningful information."[104] A Digital Object would normally constitute preserved bit sequences (or bitstreams) while the Representation Information would include the relevant software which can interpret and render it in a meaningful way. Ideally, this would also pertain to any relevant documentation and information as well as preservation and technical metadata. If format migration is the sole method of preservation it would be necessary to include the methods, such as compressing files for transfer to a newer instance of the same media. This would require updating the Representation Information to include the relevant compression algorithm due to the possibility of altering the original bits of the Content Information.[105]

The OAIS, and by extension PREMIS, were developed with a dedicated focus on handling digital content. However, the former is a high level reference model while the other is geared towards more granular implementation which means they do not always place the same amount of emphasis upon certain practices.[106] Models such as PREMIS and the OAIS place an

[102] Lee, "Open Archival Information System," 4027.
[103] PREMIS Editorial Committee, *PREMIS Data Dictionary, Version 3.0* (June 2015): 1-2.
[104] Consultative Committee for Space Data Systems (CCSDS), Reference Model for an Open Archival Information System (OAIS) (2012): 4-20 to 21.
[105] (CCSDS), "(OAIS)," 5-8.
[106] The OAIS, for instance, does not focus upon emulation and similar methods of preserving the look and feel of software as (at the time of publication of Version 3.0) they are considered emerging technologies with many

emphasis upon describing relationships between materials and objects which can help archivists to better determine where to focus their efforts. The Information Object amalgamates archival materials, or Data Objects (physical or digital), with contextual materials (Representation Information) into a coherent, manageable unit. PREMIS provides a supportive framework to further elaborate upon OAIS frameworks.

The second significant concept is the Functional Model which serves as a roadmap for ingesting, preserving, and disseminating archival materials. Specifically, it is built around the concepts of the Submission Information Package (SIP), the Archival Information Package (AIP), and the Dissemination Information Package (DIP).[107]

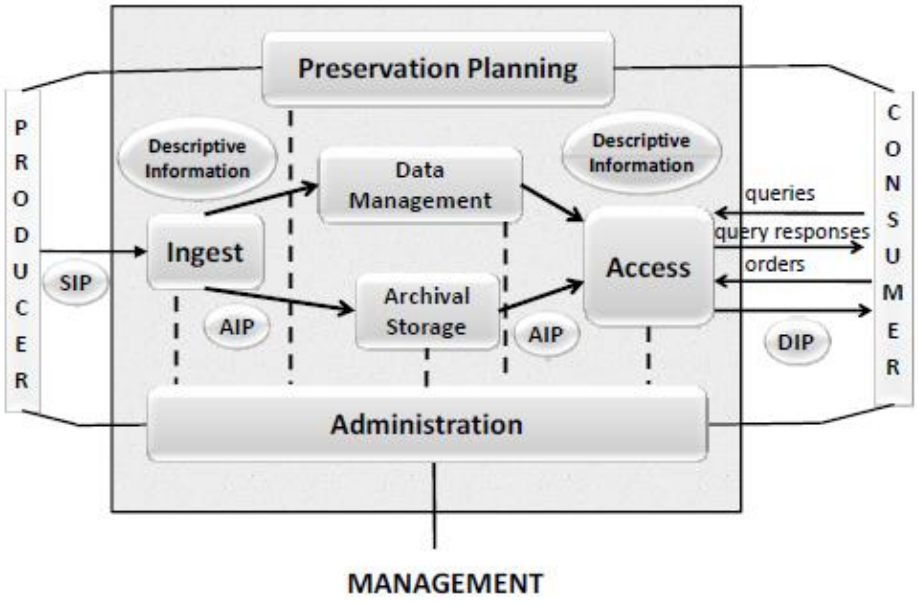

Figure 4-1: OAIS Functional Entities

From the Reference Model for an Open Archival information System (OAIS), Magenta Book (June 2010) page 4-1

---

difficulties. This paper has already examined the technical hurdles associated with emulation and while projects such as EaaS show promise. Emulation is not essential to preserve the vast majority of digital records within archival holdings. Despite this, if it is deemed necessary by the institution "emulation software logically becomes an extension of the Other Representation Information as it directly addresses Content Information preservation." Following the PREMIS model emulation would be identified as an essential preservation function under preservationLevelType as a form of "Logical Preservation" necessary for the Representation of a digital object.

[107] Brian Lavoie, *The Open Archival Information System (OAIS) Reference Model: Introductory Guide, (2nd Edition)*, Digital Preservation Coalition, DPC Technology Watch Report 14-02 (October 2014): 2. Accessed April 11, 2020, https://www.dpconline.org/docs/technology-watch-reports/1359-dpctw14-02/file.

As with the Information Object they are not individual records, because the OAIS was designed for higher level implementation, but amalgamations of content and information. SIPs are records and materials received from Producers (individuals, institutions, etc.) which are then subjected to the ingest policies (such as quality assurance) of the archive to become AIPs, meant for long-term preservation within the OAIS model, while a DIP is created deriving materials from one or more AIPs and passed along to the consumer.[108] The AIP is not a static object within the Functional Model, nor should it be, since it represents collections of records and materials within archival holdings which are subject to changes such as the addition of new material and information. The AIP is made up of Content Information and Preservation Description Information to form an Archival Information Package. Packaging Information allows the AIP to be identified and located as a single logical unit and Descriptive Information supports discovery and dissemination of the package.[109] An AIP also includes original bit-for-bit copies of digital content as well as the preservation copies. Once an AIP is created and properly secured within archival storage it is then a matter of making content available to users in the form of a DIP.[110]

As stated above, the OAIS is a reference model without a granular focus. It is meant to describe relationships and associated concepts. The act of implementation is beyond the scope of what it was designed to accomplish. "The goal is instead to provide a conceptual model of the information objects managed by an OAIS-type archive. Implementation of these concepts will

[108] Lee, "Open Archival Information System," 4025.
[109] Lavoie, The Open Archival Information System, 18.
[110] Once the appropriate content for a DIP have been determined they are not accessible without some sort of interface through which they can be delivered. Otherwise, the content remains on a server and no request can be made to initiate the creation of a DIP. Access to Memory (AtoM) developed by Artefactual, in association with the International Council on Archives (ICA), is a web-based open source application dedicated to the development standards-based archival description. In this sense it is a front-end solution to the lack of direction within the OAIS regarding implementation. It is not a source for descriptive information, only a means for archivists to publish descriptive information utilizing supported templates. It also provides adopters with the abilities to crosswalk metadata between standards and export archival descriptions and authority records for use outside of the AtoM work environment.

depend on the specific architectures, systems, and schema employed in a particular archival setting."[111] While the adaptability of the OAIS enables widespread adoption across multiple professional fields implementation is dependent upon adopters already having the necessary equipment and systems in place beforehand. Essentially, potential adopters must prepare themselves by implementing descriptive standards and the relevant tools to build upon the high level concepts and frameworks describe within the OAIS.

Closing

Acquiring documentation and making it available, and developing descriptions are not straightforward processes. Often dedicated institutions such as the Internet Archive and the Computer History Museum alleviate the burden of locating and storing such valuable documentation for archivists wishing to supplement existing digital records. However, the tangible value to users will be very limited beyond the most enthusiastic and those already well-versed in the technical side of the histories of computing. Alternative documentation affords archivists with more engaging materials ranging from official documentation to modern-day hobbyist efforts. Admittedly relying on social media sources such as YouTube may be somewhat precarious since whoever controls the channel controls whether or not the content remains uploaded. However, the fact still remains that while supplementary documentation can provide greater context regarding the record-environment relationship, such material is in no way a replacement for digital environments.

Understanding the hurdles presented by numerous non-universal standards is significant because it will in turn influence how effectively environments, or original and alternative documentation can be integrated within archival description. The fact that "standards" and "compliance" do not carry the definitive weight of technical specifications means that archivists

[111] Ibid, 18-9.

have the flexibility to adjust them as need be. It is also a long journey to develop rich, accessible descriptions due to the need to navigate multiple models, standards, and adapt them to the resources and policies available within a given institution.

Conclusions

Digital preservation, and of environments in particular, present numerous hurdles that cannot be solved by the adoption of a single method and each in turn present their own unique challenges. There are also a variety of solutions, many with the backing of various parties and institutions, dedicated to making the preservation of digital objects (including environments) more practical and comprehensive. I have repeatedly stated throughout this thesis that the aim is not to dispel the use of format migration as a method of preservation. Rather, that archives should move towards incorporating alternative methods of preservation to address the absence of environments to better leverage the functionality of digital records and provide users with a more immersive experience.

Outside of dedicated museums, acquiring and preserving complete legacy systems is not practical in the long-term. Legacy hardware is essential for digital preservation since it enables archivists to access media and rescue at-risk data. Current applications of legacy hardware in digital preservation efforts also serve as a model going forward. However, the fragility and ever decreasing availability of essential peripherals, such as floppy disk drives, does highlight the fact that there is a need to identify significant contemporary systems and components in preparation for the day when they will become essential legacy equipment. Early identification and implementation of modern environments into workflows before they become obsolete will also provide younger archivists with firsthand exposure to systems and software that will one day fall out of regular use but prove to be essential to their work.

Emulation is by far the most promising method of preservation and has seen significant developments geared towards archives and preservation efforts. While it may also be the most difficult method to undertake it has the capacity to provide the greatest results. The reproduction of digital environments will enable users to directly interact with file and folder structures developed by individuals and institutions in a more authentic manner. Through such direct interaction, users will be able to better grasp significant details and the overall structure the interconnected systems of environments that are comprised of recordkeeping and computer systems. In turn, they will be able to better understand, to some extent, the agency of records creators afforded by the environments available to them. The move towards automation and cloud-based service offer archives, anyone in fact, the opportunity to access emulation and adjust it to meet their needs. Automation and the centralization of resources reduce the technical burden on archives and afford users the opportunity to experience and, hopefully, better understand and appreciate record-environment functionality.

Description and documentation are the most limited since they cannot preserve functionality, but are also the easiest to undertake. It requires no significant investment from a technical standpoint and can easily be adapted to meet the mandate of any archive. The greatest challenge is figuring out exactly where environments fit within archival description. Some descriptive and metadata standards have begun to recognise that environments are significant cultural and digital objects in their own right (to varying degrees). However, it is still up to archivists to determine how to incorporate environments – conceptually if not functionally – through description and supplementary documentation.

Unlike paper records digital records developed with the assumption that manipulation would be an integral part of their purpose and functionality. Through rendering and interfaces

provided by environments this affords an unprecedented level of interaction between the user and the record. A digital record consists of more than its contents. It is only truly complete when rendered properly to produce an authentic performance. To better provide access to or sufficiently describe this experience will enable users to obtain greater insight into the agency afforded to the creators of records. It is this agency, after all, that has driven the development of the record-environment relationship and influenced the role of digital environments within the working and public spheres.

Chapter Three – The William O. Pruitt Jr. fonds and Digital Strategies

This thesis has discussed four different methods of preservation with respect to digital environments including their strengths and weaknesses in this regard: format migration; preservation of original systems; emulation; and description and supplementary documentation. Chapter Three will be a case study to see how these methods could potentially be applied to existing legacy records and media from a fonds within the holdings of the University of Manitoba Archives & Special Collections (UMASC). The case study will focus upon the legacy media and digital records in the William O. Pruitt Jr. fonds. The case study will also include comments and concerns raised by two professional archivists from the University of Manitoba's Libraries Research Services & Digital Strategies team (RSDS) in response to a questionnaire regarding the methods and tools they utilise in their preservation efforts.[1] The purpose is to include critical real-world perspectives, where appropriate, regarding the methods and resources discussed in this paper. The Pruitt fonds was selected based upon two criteria. First, was that it had already been processed by UMASC archivists and deemed to have archival value according to their policy. The purpose of this case study is not to question or critique the archival evaluation process or archival theory; archival value is assumed. The purpose is to demonstrate how this value may be better leveraged through the methods of preservation discussed in Chapter Two. The second criteria was that fonds was already fully described, with the media containing the electronic records described in detail. Additionally, the floppy disks had descriptions of files, formats, and compatible software written on labels so that it was possible to derive a clear

[1] The questions and full anonymized responses can be found in Appendix A and B following this chapter. Pseudonyms of "Participant A" and "Participant B" will be utilized to note direct or indirect quotations utilized throughout Chapter Three. As per the University of Manitoba Research Ethics Board's (REB) requirements both Participants were provided with the option to remain anonymous and have chosen this option. Hence, the utilization of pseudonyms. The responses provided by both participants reflect their professional opinions within the context of their jobs and personal experience in the field of archiving. They may be viewed in full the appendices following this chapter.

understanding of what exactly would be subjected to the digital preservation process and how the methods discussed could be implemented. UMASC and RSDS operate within the current consensus on format management and migration as methods of preserving digital records and materials. As this thesis has argued for the inclusion of alternative methods in addition to format migration it was important to both analyse a fonds and read firsthand accounts of professional archivists operating within this context.

William O. Pruitt fonds Electronic Records

The Pruitt fonds contains a variety of research materials, notes, correspondence, and other works published and collected by Dr. William Obadiah Pruitt over nearly 30 years teaching at several universities – including the University of Manitoba – and working the Taiga Biological Station.[2] Dr. Pruitt's area of study was natural history, specifically conservation and snow ecology, which included the launch of Taiga in 1973, initially out of his own pocket, in order to study plant and animal life in the boreal forest of North America and Russia.[3] The portion of the Pruitt Fonds that is the focal point for this chapter is the content from the section titled "Electronic Records 1988-1996" and contains a variety of legacy media and file formats pertaining to Dr. Pruitt's work.[4] Currently all digital content remains stored solely on the media that was part of the original acquisition and therefore could potentially benefit from all of the methods of preservation discussed in the previous chapter.

Almost all of the storage media in the Pruitt fonds is labelled with lists of the contents, and in some cases brief instructions on how to install the necessary included application

[2] "William O. Pruitt Jr. fonds," University of Manitoba Libraries and Special Collections, Accessed July 28, 2019, https://main.lib.umanitoba.ca/william-o-pruitt-fonds.
[3] "Dr. William O. Pruitt, Jr.," Winnipeg Free Press Passages, Accessed July 28, 2019, https://passages.winnipegfreepress.com/passage-details/id-158565/PRUITT_WILLIAM.
[4] University of Manitoba Archives and Special Collections, "William O. Pruitt an Inventory of His Papers at the University of Manitoba Archives & Special Collections," Last Modified April 22, 2014, http://www.umanitoba.ca/libraries/units/archives/collections/complete_holdings/ead/html/pruitt_10_112.shtml#series10.

software. This offers a limited understanding of the potential preservation needs of some of the records before engaging in data recovery. All the disks appear to be in good condition, some even residing in sleeves or cases, apart from fading labels. The electronic records in the Pruitt fonds includes a variety of modern and legacy media, such as CD's, cassette tapes, and 3.5- and 5.25-inch floppy disks, dated between 1988 to 1996.[5] There are also several smaller square tapes labeled Stenorette indicating they were made for use by dictation machines, specifically those produced by Grundig Business Systems.[6] As stated earlier many of the disks have dates, content descriptions, and operating instructions on their labels and cases. Some are also grouped together as part of a larger project. This holds the potential benefit that, should any of the media be inaccessible or contents unreadable, the archivists would at least be able to cross reference the non-digital records to see if any potential substitute exists.

While there are a variety of formats several feature more predominantly than others. WordPerfect and Microsoft Word (both often specified for DOS, Windows, and one for Apple) file formats are usually identified by name, including versions, though sometimes the extensions ".wpd", ".wp" or ".doc" are used.[7] One of the most convenient aspects is the frequent inclusion of the software names and versions required to access the documents on individual disks such as "WORD PERFECT 5.1 for MS DOS" and "Word for MacIntosh 5.1" for example. Other cases are vague, such as the note stating, "Terminal Emulation Software", which is unhelpful since if it does not refer to an actual copy existing on the disk there is no way of determining what specific software it is referring to. While there could be mention in the paper records this would mean dedicating a potentially considerable amount of time and effort to review the rest of the fonds.

[5] University of Manitoba Archives and Special Collections, William O. Pruitt Jr. fonds, "Electronic Records 1988-1996," accession number EL 25 PC 28, Box 1-2.
[6] "Our History," Grundig Business Systems, Accessed July 30, 2019, https://www.grundig-gbs.com/index.php?id=483&L=1.
[7] In the early days of desktop computing there was no standardisation regarding file extensions so users could use whatever they liked.

Some files are mentioned in name only and give no hint to the specific format they are in such as several floppies simply labelled "RIVMAP" numbered one through thirteen which seem to involve the gradual melting and retreat of Lake Agassiz. If these are image files they will have to be viewed or copied in order to determine their exact specifications.

The naming conventions are not always straight forward. Often it appears a form of shorthand was used and the combinations of letters and numbers do not hint in any way what data the specific file may contain. To develop descriptions at the file level, which is a difficult task for any archive due to the sheer volume of records and materials, this poses an added degree of difficulty, particularly if the necessary equipment is not available. However, there are still ways in which the data on the disks can be retrieved and preserved following the methods discussed in the previous chapter.

Preservation

Data Retrieval and Ingestion

Regardless of the form of preservation employed the first step is to attempt to retrieve the data from the disks so bit-for-bit copies can be stored on more modern and secure media within UMASC holdings. Retrieval and ingest will also require using tools and methods that are foundational for all four methods of preservation so they will briefly be mentioned here and explored in greater depth in the sections to follow.

The fact that the majority of legacy media are 3.5-inch and 5.25-inch floppy disks, means that legacy hardware will be essential for data retrieval. Fortunately, the Digital Strategies team has access to tools for reading legacy media, specifically a recently acquired Forensic Recovery of Evidence Device (FRED).[8] As covered in Chapter Two, a FRED is capable of bit-level duplication of digital media content while ensuring no alterations are made through the use of

[8] Vielfaure, "Have you Heard About FRED?"

write blockers. This means the data can be preserved on more stable media, readily accessible for format migration or unforeseen future projects, such as emulation. The capability of connecting with a wide range of peripherals in order to read legacy media was also an important factor for Digital Strategies. This meant "It would make the process of working through the backlog of digital material more efficient as the alternative would be to brainstorm a new way forward every time we encountered new types of media that our standard workstations couldn't support."[9] Ingest will therefore not be an issue for the material in the Pruitt Electronic Records with regards to accessing legacy media.

Format Migration

Each of the alternatives to format migration are not viable as stand-alone methods of preservation. The fragility of legacy systems makes them impractical for frequent use. Even if Emulation-as-a-Service (EaaS) attains enough widespread adoption to become a standard fixture in many archives most users will still desire, or demand, to access preserved data in more timely and simplified manner. Documentation and description are pointless if there are no digital records to begin with. To reiterate, while current applications of format migration as a method of preservation are focused upon making content less dependent upon original environments this does not make it an invalid method of preservation. Rather, it should be used in combination with one or more other methods to enable the inclusion of environments.

While the labels on the disks from the Pruitt fonds make it possible to form reasonable assumptions as to what formats are on many of the discs the point of data recovery is to use the appropriate tools and resources to be certain. Format migration is only possible if the format of the original record can be identified, otherwise it is impossible to determine the migration path and potential target format(s). Developing a proper AIP (Archival Information Package) also

[9] Participant A, in response to Question Five.

requires more than the preservation and migrated versions alone, it is important to include the relevant Preservation Description Information (PDI) metadata addressing the migration. One useful resource is Archivematica, developed by Artefactual (who also developed AtoM[10]) as "an integrated suite of free and open-source tools that allows users to process digital objects from ingest to archival storage and access in compliance with the Open Archival Information System."[11] Whereas AtoM is a front-end solution and focuses upon description and descriptive metadata Archivematica is a back-end solution focused on SIP (Submission information Package) to DIP (Dissemination Information Package) development and maintenance.[12] Among the numerous functions it can perform are identifying formats, generating metadata following international standards, and performing format migration, all of which can be automated.[13] Some of the identification tools found in Archivematica utilize the PRONOM technical registry to aid in identifying file formats.

Since archivists have access to such a robust infrastructure of tools and resources dedicated to format migration it is understandable why this is the go-to method of preservation for at-risk data. There is also the fact that format migration is meant to move away from technical dependencies that are difficult or impossible to maintain in order to streamline the process of preservation. "Through migration, we have a more controlled number of formats and software to manage and monitor. It allows us to better manage changes over time as technology and record creation trends evolve. And ultimately, it takes a lot less resources in terms of time,

[10] Access to Memory.
[11] "What is Archivematica?," Artefactual Inc., Accessed April 15, 2020, https://www.archivematica.org/en/docs/archivematica-1.11/getting-started/overview/intro/#intro.
[12] SIPs being the materials received from Producers (individuals, etc.), AIPs collections of ingested and additional relevant materials for long-term preservation, and DIPs derived from one or more AIPs for dissemination to users.
[13] Bronwen Sprout and Mark Jordan, "Archivematica As a Service: COPPUL's Shared Digital Preservation Platform," *Canadian Journal of Information and Library Science* 39 no. 2 (2015): 240.

staff and finances."[14] Migrating the Pruitt fonds will not be difficult for the majority of records, assuming they are mostly word processing, spreadsheet, and image files. Since many of the records correspond with published materials there is no reason to necessarily upload access copies for immediate access. It would be more reasonable to maintain them to be provided upon request which will be relatively simple as there are no limitations placed upon access to the materials within the Pruitt fonds.[15] There is also the fact that the originals remain available as part of the AIP so that records that cannot be migrated may still be accessible through some other means in the future. As Participant B stated, "we are thinking of future archivists who might want to perform actions on the object that we might not have envisioned."[16]

Accessibility via format migration may not be uniform across the whole of the Pruitt digital records. Some files may be difficult or impossible to migrate while others may lose their original functionality. While images and even files created in word processors may retain all of their content upon migration more interactive ones such as spreadsheets and multi-media presentations could lose their underlying functionality. There is also the concern that content will not be properly rendered due to differing font and symbol lexicons and methods for handling alignment between the original and target formats. In more severe cases content can also simply be lost or rearranged to the point that the resulting version is illegible. Digital records are inherently more complex than paper and "Given the amount of code used to render data graphically, it is difficult to capture every detail of a record's functionality, as well as a faithful legacy representation, with any digital preservation approach."[17] There is also the fact that unless time and effort is spent checking every single record it would be impossible to determine the true

[14] Participant A, in response to Question One.
[15] University of Manitoba Archives and Special Collections, "William O. Pruitt."
[16] Participant B, in response to Question One.
[17] Participant B, in response to Question Seven.

extent of the differences between the original and migrated versions. The majority of files in the Pruitt fonds are from WordPerfect and Word so the most significant loss will be in functionality since access copies will be in more static formats which will mean loss of access to the editing functions of the original applications software.

There will also likely be difficulties accessing some of the material, such as the two atlases that are part of the "Windows on Russia" group, since the labelling indicates it is in fact a program saved across multiple disks. While it may be possible to extract some individual files containing text or images without the program they will be fragments without context or functionality. Geoffrey Yeo states that "Collections, observably, have part-whole structures; they are assemblages of items, and each item in the collection is acknowledged as a member of the whole."[18] He elaborates that "A further characteristic of sub-collections is that they can be seen both as divisions of the larger whole and as collections in their own right."[19] This would encompass the "Windows on Russia" record group, not only as a part of the larger Pruitt digital records but the multiple disks that together make up the program. The question then is how much, or how little, can be left behind and changed via format migration before it becomes detrimental in some manner to the immediate records and larger Pruitt fonds?

Original Systems

In some ways utilizing original systems would be the most difficult preservation method to apply to the Pruitt fonds. Many modern digital preservation tools have ongoing support from developers such as Digital Intelligence which offers training in the implementation and use of their products such as the FRED.[20] Some, like Archivematica and AtoM, also have very

[18] Geoffrey Yeo, "Bringing Things Together: Aggregate Records in a Digital Age," *Archivaria* 74 (2012): 50.
[19] Yeo, "Bringing Things Together," 50.
[20] "Overview," Training, Digital Intelligence, Accessed April 20, 2020, https://digitalintelligence.com/training/overview.

dedicated userbases that serve as another source of support. Legacy environments do not have such extensive resources, however. While there are dedicated groups and efforts around preserving older computing systems both in the form of institutions such as computer museums or modern hobbyist movements. But with no production to replace damaged or lost components a legacy system is arguably more in need of preservation than the records they are intended to render. This is made all the more difficult due to the fact that no form of comprehensive coverage exists for obsolete systems since some attract a user base while others see little interest.

Components of legacy systems are already a vital part of digital preservation programs due to the simple fact that "We can't read the media without the right hardware, so it plays a big role in preserving legacy media. If we can't read and extract the content, we can't preserve it."[21] There is also the fact that records creators and donors are ultimately in control of how records are formatted and organised (or not) when they are sent to an archive. "In reality, archivists rely on creators to bring their data up to contemporary formats before transferring it to archives, but given that the traditional approach to acquisition takes place some years or decades after those records were originally created, there is often a huge technology gap to be bridged."[22] The gap between creation and acquisition is not the only factor. Existing backlog can also exacerbate the technology gap. Records could be submitted today in relatively safe formats but can spend years on a shelf before archivists manage to recover the data.

Fragility and scarcity make legacy hardware a limited and diminishing resource, one which archivists must also remain diligent in monitoring:

> The main challenges relate to the age of the media and the hardware. The longer the media has been obsolete, the harder it can be to find the hardware to read it. Furthermore, that hardware will not remain operational forever, so we need to

[21] Participant A, in response to Question Four.
[22] Participant B, in response to Question Four

> monitor the hardware, just like we monitor the files, to make sure that we can refresh it as needed to continue and read legacy media as it comes in.[23]

These challenges will only be exacerbated by the adoption of full legacy systems beyond the peripherals in use. The benefit of utilizing an intact legacy system is that once you have it up and running you are all set, for the most part, from a technical standpoint. All that is required is to load the media with the requested records and move on. Anyone who enters the archive can access the system so long as they obey the necessary rules which include no foods and avoiding physical damage. Digital specific rules would be not to modify the system, content, or media. Since the Pruitt Electronic Records are not bound to a specific personal system there is a degree of flexibility regarding what systems should be acquired. The label descriptions indicate hardware running later versions of MS-DOS, Apple-DOS, and Windows 95, with the necessary applications software of course, would be suitable and ideally also contain the appropriate disk drives so that Digital Strategies are not deprived of the ones necessary for their work. Such generalisation would also make these systems useful for other legacy media and records collections with compatible content so the benefits extend beyond Pruitt. Such systems would most likely be located within a reading room with scheduled use to limit wear and tear. Unfortunately, many of these benefits also bring up points of concern, both immediate and long-term.

Galloway argues that when it comes to preserving older digital records "we need to be able not only to read and copy them, but also to recover the performance of the environments themselves in order to understand the environment in which the creation was done."[24] One question this poses is how to determine what an authentic performance actually is. If it is simply adhering to the specifications provided by the labels, and further detailed through metadata

[23] Participant A, in response to Question Four.
[24] Galloway, "Retrocomputing," 624.

capture when creating preservation copies, then any generic emulation of MS-DOS with WordPerfect will be sufficient. The last part about "the environment in which the creation was done" creates an issue, since there is no way of knowing the specifications of the computer systems involved. There were no computers included in the donation. There are also the matters of acquiring and maintaining compatible systems and the issue of human interaction. Owners of fully preserved or restored legacy systems are unlikely to part with them easily so it is more likely that some degree of work would have to be done on whatever hardware archive would be able to get a hold of. It is also difficult to tell how long a system will remain relatively sound overall before software and components begin to fail and must be refreshed. The alternative would be to construct a system but this would be an overly complicated process more in line with the Archives of Ontario addressing a severe limitation in their preservation capabilities. Both scenarios would essentially lead to "Frankenstein" machines that exist more as technologically primitive analogs of a modern FRED since they exist to retrieve and access data rather than enable users to create, render, and manipulate it. This makes them increasingly incapable of reproducing an authentic performance. While modifying and adapting legacy systems is not some form of heresy, it has in fact proven essential in some of the cases discussed in this paper, it highlights the fact that it is dubious to rely upon such equipment as the sole means of reproducing authentic performances.

Then there is also the issue of the knowledge, "practical but undocumented as well as tacit, which make it possible to perform the task of preserving the performance artifacts, hardware, and software."[25] The majority of users that enter an archive will have little to no experience interacting with legacy environments, particularly millennials who will most certainly not have this tacit knowledge, save for enthusiasts. Unless the content is of significant enough

[25] Ibid, 632.

value (historical, statistical, etc.) and potential user interest to denote the need, it is simply not practical to utilize a legacy system in such a limited capacity. The Pruitt digital records alone do not seem to merit this.

Emulation

There are multiple options for implementing emulation with regards to the Pruitt digital records (and ideally for the benefit of other fonds). One immediate benefit is that the University uses fairly standardised general use workstations available to students and visitors across campus so any compatible emulator could be accessed by any of the public workstations within the UMASC reading room. The potential would also exist to make the emulated environments available across campus, with restrictions imposed upon records where appropriate. While the Pruitt fonds is the primary focus of this case study any emulation project would benefit from, and likely require, the capacity for broader practical application. Ultimately the goal is twofold: provide users access to digital records in an authentic reproduction of a legacy environment; and, avoiding harming or exacerbating existing physical or bit-level damage to the original media and record versions.

Whether developing the necessary emulated environments in-house or relying upon external services it is essential to ensure that the necessary information is preserved along with the records before-hand. Tools such as PRONOM are useful for identifying file formats but they "cannot identify or record the dependencies that specify the emulator, operating system and application needed to emulate a preserved artefact."[26] Emulation, as Rosenthal has discussed, is a form of mimicry, and an incredibly detailed form at that. All with the aim to reproduce an authentic performance of records in their original formats. The File Information Tool Set (FITS) "is a free and open source tool for identifying and validating file formats, extracting metadata

[26] Rosenthal, "Emulation & Virtualization as Preservation Strategies," 14.

embedded within files, and outputting the metadata in various formats."[27] If the mimicry of emulation is to be successful it will require acquiring and preserving as much detailed information as possible to prevent miscommunication between the layered programs and relevant software. Data corruption at the bit level is also an ever-present risk and can make records inaccessible regardless of the quality of the emulator. In practice multiple copies must be kept through replication[28] of data on multiple types of storage and they must be audited (scrubbed) to detect and repair damage, but the costs of doing so limit how frequently this can be done.[29] Upon capture and ingestion the Pruitt digital records might have a relatively small storage footprint but they add to the already significant volume of material under the custody of UMASC. Both the fragility of digital media and content, and the costs related to preservation would affect the allocation of resources to an emulation project.

It is clear that a browser-based user interface would be the best choice for user accessibility since it is already a standard method utilized by the examples discussed earlier in this thesis. While interacting with the emulator itself may seen otherworldly to some users, especially without the tacit knowledge, web browsers are familiar. The in-house option also means that the University, UMASC, and RSDS are in complete control of the records and environments. Limiting use to a specific browser also makes development and maintenance easier since compatibility only needs to be maintained for the one browser rather than multiple. The Internet Archive's efforts take a different approach by requiring the emulator to be temporarily downloaded and run using the browser (specifically the JavaScript engine) which is

[27] "FAQ," File Information Tool Set (FITS), Accessed April, 22, 2020, https://projects.iq.harvard.edu/fits/fitsfaq.
[28] A backup is a copy of data meant for long-term storage while replication involves the frequent creation of copies overriding older ones, usually for the purposes of rapidly replacing recently lost and/or corrupted data. "Storage 101: Replication vs Backup, an Synchronous vs Asynchronous," Computer Weekly, Accessed June 19, 2020, https://www.computerweekly.com/feature/Storage-101-Replication-vs-backup-and-synchronous-vs-asynchronous.
[29] David S. Rosenthal, "Bit Preservation: A Problem Solved?," *The International Journal of Digital Curation* 1 no. 5 (2010): 142-3.

possible because the software and operating systems were originally designed to be run by computers with significantly less processing power.[30] This becomes less practical for operating systems and programs from the 1990s onward which grow from a few hundred megabytes to several gigabytes in size. For the Pruitt fonds size is fortunately not an issue if this was the chosen development path as the operating systems and files are from the DOS and early Windows eras. This would circumvent compatibility issues with physical systems when they are replaced as time goes on and allows access for users outside the archive. If this degree of accessibility is considered too much then it is simply a matter of limiting the emulation to select workstations within the UMASC reading room. Developing a compatible in-house emulation solution will mean finding people with the skills and knowledge to develop new or adapt existing emulators and maintain them.

One solution then is a cloud-based EaaS option like the Baden-Württemberg Functional Long-Term Archiving and Access (bwFLA). One of the reasons for choosing the open-source Archivematica was that it "allowed us to maintain all of our data intact and independently from a vendor. It gave us something scalable, and gave us control over where our data was stored to ensure that it could be stored on local/Canadian servers."[31] By utilizing bwFLA nothing needs to be installed on University of Manitoba computers and data remains on the same servers. Processing power and many other technical requirements are not a concern, aside from making the appropriate records accessible. The overlay approach taken by bwFLA means that it is a simple matter to request an environment that meets the dependency requirements of the material from the Pruitt fonds; it is simply a matter of selecting the appropriate operating system and software version (MS-DOS with WordPerfect version 5.1 or 6.0 for example). Another benefit of

[30] Rosenthal, "Emulation & Virtualization as Preservation Strategies," 9.
[31] Participant A, in response to Question Eight.

bwFLA is that it can generate "use" metadata which can be used to justify whether or not to continue funding subscribing to an EaaS and potentially expand to include future fonds, or that it is not worth pursuing.[32] The recent development allowing users to download and run bwFLA locally rather than an external cloud also places more control in the hands of RSDS. This also helps to negate some of the concern over utilizing a third-party tool should development abruptly end. While RSDS would "either have to find a way of replacing these emulators, and do so quickly, or abandon emulation entirely and find an alternative" maintaining a local copy of bwFLA would provide some breathing room for an exit strategy. "That exit strategy should ensure that we can abandon these systems at any time without losing significant resources or efforts, which we couldn't do if we relied on these tools."[33] In the proposed scenario for this case study emulation is an extension of the current format migration based program, which ensures access copies still exist in alternative formats. The emulators are also still within their possession including all the relevant data created and collected over time. So if support for bwFLA were to end abruptly nothing is irrevocably lost.

While EaaS does remove the burden of development, processing power, and maintenance from its users the fact still remains that the University has made an investment in this service. Aside from the obvious financial investment there are other requirements on the user's end. First there is the task preparing records and collecting the necessary metadata to ensure identification of records and communication between systems. Staff must also familiarise themselves with the software in order to make sure it functions properly and assist visitors and clients and there is the need to incorporate emulation into existing workflows. One of the most significant determinates will be whether users of archives will even utilize EaaS to access the records to any significant

[32] Rosenthal, "Emulation & Virtualization as Preservation Strategies," 15.
[33] Participant A, in response to Question Two.

extent; “I wonder how important the “functionality” of older digital records is for users. Presently, emulators are nice to have, but not a have-to-have to represent the information contained within the data.”[34]

Documentation and Description

Description and documentation is the least technically intensive alternative to supplement format migration, but also the most limited since it does not replace digital environments, merely includes more detailed references to them in an effort to better establish the record-environment relationship for users of archives. If an archive does not pursue a means of direct environment preservation this is the simplest method to adopt since it mostly builds off of preservation practices utilized for records during regular ingestion and format migration. Some of the descriptions on the labels of the disks in the Pruitt digital records can at the very least be included in the finding aid. This would be the bare minimum effort with the records remaining as they are, solely on the original media. The scenario of this case study assumes that ingestion, data retrieval, and format migration have been performed so there will be more detailed technical information available thanks to the resources discussed earlier in this paper.

The Pruitt fonds does not include item level description for paper or digital records but the potential for including information regarding format migration does exist in a form of documentation generated during the data recovery process.

> As we preserve the content on behalf of the Libraries’ archives department, they would be responsible for the descriptive metadata, so I can’t speak to how they would accommodate this information, but generally, archival description standards don’t really accommodate for that type of information very well. The PREMIS metadata we generate and preserve is better suited for this purpose. If users were interested in knowing what transformations a given record went through, this would be what we would look at to answer those questions.[35]

[34] Participant B, in response to Question Two.
[35] Participant A, in response to Question Six.

While incorporating this level of detailed information directly into the finding aids and descriptions would be difficult it can be referenced and made available upon request as a separate DIP. Of course archival description is meant to communicate to users of archives the essential information about the records and materials they may wish to view. On the matter of archival description, Participant B did note that "I am somewhat removed from the creation of archival description in my current duties. That being said, the migration logs generated by software such as Archivematica do record these types of actions as could/should be included within an archival description."[36] While not ideal, AtoM, the archival description system employed by UMASC, does include the ability to upload digital objects so this metadata could be made immediately available for users. This method could extend to include spreadsheets containing more detailed information about the disk images generated by Digital Strategies.[37] Alternative supplementary materials would include the *Getting To Know: WordPerfect 6.0 for DOS - VHS (1993)* and uploads of *Computer Chronicles* mentioned earlier as well as links to Internet Archive emulated software and operating systems.

Despite all this it is difficult to convey to potential users what functionality may be lost through conversion. Participant A notes that "While my work doesn't involve creating finding aids for this content, something like a Physical Condition field, or a Conservation Note could accommodate this information."[38] Such adaptations will initially be time consuming to establish within current descriptive practices but once this is accomplished it can be replicated across multiple fonds as an internal standard. This will still require the effort to implement though before attending to existing descriptions. A simpler starting place would be at a higher overarching descriptive level above even the fonds level; "Sometime in the near future, we'd also like to include a format

[36] Participant B, in response to Question Six.
[37] Participant A, in response to Question Eight.
[38] Participant A, in response to Question Seven.

policy registry that lists the preservation and access formats we maintain on a public-facing webpage, so that users can get a more general sense of the transformations that take place as well."[39] While description and documentation should ideally focus on specific record-environment relationships, utilizing a broader method to communicate the existence of such relationships can still help archivists to educate users about the complexity of materials they wish to view.

Conclusions

While the inclusion of digital environments and supplementary materials with the Pruitt digital records is possible using each of the methods discussed in this thesis some are better than others. Relying upon format migration alone will still produce important technical information that can potentially be included within current archival description as supplementary material. The issue is where exactly this information fits within archival description. The development of new fields of description would almost certainly be required to accommodate such information beyond simply an inclusion in the notes section. This will have consequences beyond the Pruitt fonds since it will mean establishing a new in-house standard which will need to be applied to other fonds and collections or else it will be a one-off and an oddity. The inclusion of this and other supplementary material will aid in making it discoverable to users and while not directly part of archival description will still be available to users. This would – hypothetically, still mean there is no preservation at the environment level though.

The preservation of original hardware beyond what is already used for data recovery and the development of disc images is the most impractical in both the short and long-term due to the need to first locate one or more intact and functional systems and to maintain them. The fact remains that they will almost certainly see little use, not only because of limited user interest but

[39] Participant A, in response to Questions Six.

also because access will have to be limited to small intervals in an effort to limit wear and tear. There is also the apparent possibility that any malfunction, either hardware or software based in origin, would mean the end of the system's life. If not, then the ongoing need to replace and modify would eventually lead to systems that bear little in common with their original specifications so the question would then be whether the experience they afford is authentic. It could also mean peripherals which could prove essential to data retrieval would be tied up in systems that see little use.

Emulation, specifically utilizing EaaS, is the most practical route from a technical standpoint since much of the heavy lifting – ongoing development and computing power – are handled by third parties. Making this service discoverable is not difficult since it can easily be referred to in the Scope and Contents in the finding aid for the fonds. If deemed successful more legacy digital records can be incorporated at which point it would make sense to advertise the service on the UMASC and UM Digital Collections home pages. The relative similarity of record formats (for the most part) in the Pruitt fonds mean that it would be a relatively straightforward test run of EaaS. As Pruitt did not donate any original systems there are no unique user settings or related tweaks required, simply selection of relevant operating system and relevant applications software. Cost and resources are a limiting factor in preservation and digital environments are no exception. Justifying EaaS may require expanding beyond the Pruitt fonds to offer potential users greater variety in effort to show off its potential.

Going to such lengths to preserve the functionality of the records in the Pruitt fonds does not serve researchers interested solely in Pruitt's work alone. Any students or researchers focused upon the histories of computing would benefit from being able to experience such records in their original formats, regardless of the contents, and their original functionality

through an emulated legacy environment. The Pruitt fonds is a starting point. Going a step further and applying EaaS more broadly across multiple fonds would make for a broad pool of resources for researchers and students wishing to approach such material from the digital perspective. While this would first require actually applying EaaS to the Pruitt fonds to see how well it works, if successful, it would allow UMASC and users to access the current untapped potential of volumes of electronic records in their holdings.

Conclusion

This thesis began by proposing two questions. The first concerns the effect of failing to preserve digital environments, while the second, addresses how might we go about preserving them. The answer to the first is a loss of technical and social context and functionality. This thesis has shown digital environments are cultural objects that have had a significant impact upon people's public and working lives. Computer systems have also been influenced by society in turn from anthropomorphising, to being declared as symbols of modernity, to finally being ever-present in our lives. They have been integrated into recordkeeping systems, starting with tabulation machines shared amongst multiple departments within an organization and culminating in individual desktop workstations. This development was the result of technological advancement driven by the perceived and actual needs of users, as interpreted by hardware and software developers. It is this human agency which transforms the computer from being solely a technological artifact to being a cultural artifact as well.

While much of early electronic records preservation was driven by the notion of such records merely transitory or "working papers," many archivists – particularly those working in machine-readable archives – understood the significance of preserving digital functionalities for unforeseen potential use. This has persisted even through the increasing costs and concerns associated with obsolescence. Sue Gavrel argued that media transfer was essential since it ensured content remained machine-readable for future users who wished to manipulate the data without having to reproduce it. Jeff Rothenberg went a step further arguing that, given sufficient computing power, even older computing systems could be replicated through emulation. More recent voices such as Christoph Becker have argued that authenticity is dependent upon how data is manipulated to recreate a performance. As stated earlier, this means a digital record is not a

static object but a combination of content and computational processes since it must be rendered in order to be views and manipulated by the user in a meaningful way. By extension, any performance is tied directly to its environment since the environment provides the rendering capabilities. This is the essence of the record-environment relationship.

Being able to reproduce records in original formats can prove invaluable to future research. Even if an author or scientist has formally published materials the way they organise the content within their personal records will reflect their thought processes and how they came to specific conclusions including the inclusion or exclusion of content from their published works. Being able to also reproduce original functionalities down to the file level, rather than relying solely upon migrating content to different formats, affords users the opportunity to interact with preserved recordkeeping systems on a much more granular level. From here it is possible to more deeply infer how records creators and managers interacted with the broader system via the tools and environments available to them, recognizing that these environments and tools dictated much of their capabilities and limitations. This interactivity is part of the technological and social provenance of digital records and environments and cannot be fully expressed without preserving both. The significance of preserving environments goes beyond the immediate record and larger collections and fonds. Every effort to preserve and document an environment contributes to the broader histories of computing.

So how might we go about preserving digital environments? As the world continues to be reliant upon digital technologies archives will have to handle larger and more diverse volumes of diverse digital records. From a descriptive standpoint it is possible to perform basic identification and documentation of widely used computer systems and software (current and obsolete) as independent objects means that they can be linked to a wide range of digital materials.

Determining the original formats and systems allows archivists to identify hardware and software dependencies of older and obsolete media and the records stored on them. Specialised and obsolete software would require ongoing communication with donors and additional research for the development of supplementary documentation. Archives and archival theory are continually evolving and adapting to meet the needs of digital records and this is be reflected in the adoption and implementation of reference models and standards. Formal and supplementary documentation, in addition to information generated or extracted by tools during ingest, can be included within descriptive systems either as additions to current fields or through the creation and designation of new descriptive elements and fields. Archivists must look to their current descriptive methods and standards for precedent to determine where and how information regarding the record-environment relationship may be included.

Records must still go through ingest procedures regardless of the methods of preservation used. However, the inclusion of emulation means that once records are identified many may potentially be made readily available more rapidly to users rather than having to wait for mass format migration. Format migration is still essential since it is an established and efficient means of delivering content to users. But it does not leverage the full potential that exists in many records in their original formats. Emulation better reflects the record-environment relationship because it allows users to experience, to some extent, the agency of the records creators through interacting with them. Emulation-as-a-Service (EaaS) may allow for widespread adoption of emulation-as-preservation enabling users to access records in legacy formats either in reading rooms or remotely. Original file formats are always preserved to ensure authenticity in what Adrian Brown has dubbed the chain of provenance with regards to format migration. Emulation provides an additional path by allowing users to tap into the latent potential in original formats.

Resources and mandates will be significant determinants regarding how far and archives may go if they were to adopt the methods discussed in this thesis. However, as archives become increasingly inundated with digital materials it is important to find some way to better express the relationship between records and environments and the impact of human agency.

Appendix: University of Manitoba Digital Archivists' Responses to Questionnaire

Participant A

1. Why did the Digital Strategies department decide to utilize format migration as its principle method of preservation for legacy formats rather than alternatives such as emulation?

The main benefit is that it strongly reduces our software and hardware dependencies over the long-term. If we emulated our content, the types of formats and systems we'd need to have in place and support consistently over the long-term would increase exponentially overtime. We'd also have to maintain and refresh the hardware used for emulation. Through migration, we have a more controlled number of formats and software to manage and monitor. It allows us to better manage changes over time as technology and record creation trends evolve. And ultimately, it takes a lot less resources in terms of time, staff and finances.

2. Resources, such as the Internet Archive, provide access to emulators like DOSBox, licence-free early versions of Windows and Mac operating systems, and compatible legacy applications. Such tools have the potential to allow archives to provide users access to older digital records while preserving a greater degree of their original functionality. What are the concerns and limitations to utilizing such tools as a method of preservation?

A big concern is dependency on these systems. If we relied on these tools and the Internet Archive later lost interest in funding these types of projects, or could no longer support these emulators, then we'd essentially be left high and dry. We'd either have to find a way of replacing these emulators, and do so quickly, or abandon emulation entirely and find an alternative. We can't rely on these types of tools without having an exit strategy ready to go at any moment to

ensure that we can still access our content when these tools are no longer available. That exit strategy should ensure that we can abandon these systems at any time without losing significant resources or efforts, which we couldn't do if we relied on these tools.

Another limitation is that it doesn't truly preserve the look and feel of the original records, because in many cases, there isn't a single "look and feel" as the creator may have interacted with the same records in more than one environment. For example, if a donor gives you their records, including their e-mails, think about the multiple environments in which they may have accessed these e-mails. The same e-mail may have looked entirely different depending on whether they accessed it from their work computer, using Microsoft Outlook, or maybe from their phone, using an app or a mobile browser, or at home, accessing it through a browser on their personal computer. The browser access could also look entirely different if they're switching between different browsers. They might be working on a PC at work, and on a Mac at home. Overall, records can also display differently based on user preferences, or on past interactions, like how websites can display differently based on your location and what cookies are stored on a given device. News sites and social media sites are a good example of this: the same website can look entirely different from one user to the next. Whose environment is being preserved, and is it the only one they've interacted with? So while emulation might provide a user the opportunity to interact with the environment, there are several variables that can impact what that environment looks like, and it leaves archivists with a decision to make when choosing which environment they want to present to the user. With which environment should they be interacting with?

Another potential limitation is determining what you want to preserve, and whether emulation achieves that effectively. Are you looking to preserve the experience/functionality or

the content? Often, reference work is transactional: someone wants information, the archivist retrieves that information for the user and their question is answered. Other times, a user might be after an emotional connection to the records. For example, a user looking for family records might find the handwriting of an ancestor more impactful than the content of the letter itself. If our primary goal is to give users that emotional experience and connection, then emulation might be better suited, but if our main goal is to preserve the content, utilizing the description of the records to provide the context, which I would argue many archives do, then migration achieves this and uses far less resources than emulation.

3. When handling collections of mixed digital media (flash drives, DVDs, CDs, 3.5 and 5.25-inch floppy disks) how does Digital Strategies determine priority regarding what is preserved? Would newer storage media be processed first or is emphasis placed upon older, and possibly more fragile media?

We definitely consider the age of the media, but the value comes first, and will ultimately be the deciding factor. If we know that the content on a given piece of legacy media is entirely unique, or documents an under-represented group of people, or event, or information that might be entirely lost if this piece of media isn't preserved, then that would take priority, no matter what the age of the media. We'd prioritize the age of the media in cases where we have multiple pieces of media with similar value assigned to them. In that case, we'd process the oldest media first.

4. Since Digital Strategies works with legacy media, such as floppy disks, what role does legacy hardware play in preservation and what are the challenges of working with obsolete equipment?

We can't read the media without the right hardware, so it plays a big role in preserving legacy media. If we can't read and extract the content, we can't preserve it. Beyond this, the hardware needs to have the right configuration (i.e., we need to make sure that the hardware allows us to read the media with write-blockers in place so that the media isn't altered).

The main challenges relate to the age of the media and the hardware. The longer the media has been obsolete, the harder it can be to find the hardware to read it. Furthermore, that hardware will not remain operational forever, so we need to monitor the hardware, just like we monitor the files, to make sure that we can refresh it as needed to continue and read legacy media as it comes in. Overtime, more media types become obsolete and as that happens, you become more dependent on hardware that is external to your primary workstation. Workstations used to have floppy disks drives, but they don't anymore, so there's now increasing reliance on external floppy disk readers or devices like a Kryoflux to read them. Optical disk trays are soon going the same way – they're rapidly being phased out of standard workstations. As media types come and go, we increasingly rely on external hardware, so the more types of media we acquire overtime, the more hardware we need to obtain to help render it.

5. What factors were influential in the decision to acquire the Forensic Recovery of Evidence Device (FRED)?

We had a significant backlog to deal with, which was made up of a wide variety of format types and not all of them could be connected to our standard workstations, and we didn't have write-blockers in place for these workstations either. FRED gave us more options to connect various devices and media cards to a single workstation and ensure that all of these devices were write-protected upon connection. It would make the process of working through the backlog of digital material more efficient as the alternative would be to brainstorm a new way

forward every time we encountered new types of media that our standard workstations couldn't support. FRED makes our overall workflows more consistent and effective.

6. While FRED can read legacy storage media and identify legacy file formats, migration is still the method of choice for long term preservation and access. What details, such as the original environment (hardware and software) required to interact with and manipulate the original legacy formats, are included in the archival description as part of the record's provenance? (For example, noting that a record created in WordPerfect 3.0 for MS-DOS was migrated to Word or PDF format).

FRED can read legacy storage media and some legacy file formats, but it's not an emulator, and it doesn't replace migration. It's the first step towards preservation when we receive files on storage media. It allows us to access the files to get a better understanding of what they are, so that we can then move onto the preservation and access stage. For some institutions, this stage might involve emulation, but for us, it involves migration.

As we preserve the content on behalf of the Libraries' archives department, they would be responsible for the descriptive metadata, so I can't speak to how they would accommodate this information, but generally, archival description standards don't really accommodate for that type of information very well. The PREMIS metadata we generate and preserve is better suited for this purpose. If users were interested in knowing what transformations a given record went through, this would be what we would look at to answer those questions. We also typically generate a spreadsheet for disk images that will list the types of formats included in each disk image, and that would be submitted to the archives department, and they could integrate that into their descriptions as well. Sometime in the near future, we'd also like to include a format policy registry that lists the preservation and access formats we maintain on a public-facing webpage,

so that users can get a more general sense of the transformations that take place as well (e.g., if they receive a jpeg, they could infer from the format policy registry that it may have previously been a png, gif, tiff, etc., but was migrated to jpeg as an access standard).

7. Interactive media often contain embedded links, animation, interactive components, or other specific requirements in order to function (the code in individual spreadsheet cells for example). What are the limitations concerning what content can be extracted and preserved and how do you account for any lost functionality afterwards? (Are users made aware of any lost functionality in the record they are viewing)?

When possible, I think making the creators aware of the lost functionality through preservation is more important, though not always possible. Working with researchers who create data, I try and talk about some of the things they should keep in mind to keep their data usable and understandable over time. For example, if you convert an Excel workbook with multiple sheets to a csv format, you lose all but the first sheet through the conversion, so I try and encourage people, where I can, to only create single-sheet workbooks when they're using spreadsheet software. There are other cases where even if we could preserve the interactive media, we can't for legal reasons. For example, if a Powerpoint included a video that wasn't created by the donor and they don't hold the copyright, then we couldn't preserve it, whether or not we could maintain its functionality. Admittedly, that type of early interaction with the record creator isn't always possible though.

While my work doesn't involve creating finding aids for this content, something like a Physical Condition field, or a Conservation Note could accommodate this information. Ultimately, access starts first and foremost with a finding aid. To ensure people can access the

content, they need to know that it exists, so describing lost functionality there would be a good way of ensuring users have this information upfront.

8. Archivematica has a wide variety of tools for digital preservation and supports several standards. Why was Archivematica adopted by Digital Strategies and what are some of the most important features for the work you do?

Going back to the idea that for any service you use, you need an exit strategy, we wanted to avoid proprietary solutions, which don't give us a viable exit strategy. Such services might offer you the ability to exit with your data, but what version of that data? Some services might give you your objects back and the descriptive metadata with it, but you might lose all the preservation metadata and logs that the system generated and which document the changes made to the records overtime. That information is important for auditing purposes and authenticity. And whether or not they give you your data, you still lose the efforts you put into ingesting content into this system and processing it. Crosswalking whatever data you get to a new system is usually a time-consuming and messy process, so you're losing a lot of effort with other proprietary options. You also get less of a say with proprietary options - if you don't like how something works, you can't do much about it. All of these considerations are what moved us towards an open-source solution like Archivematica. As it's open source, Archivematica allowed us to maintain all of our data intact and independently from a vendor. It gave us something scalable, and gave us control over where our data was stored to ensure that it could be stored on local/Canadian servers. The automation of the service is another good benefit. It also gave us a seat at the table. More specifically, if we don't like how something works, we have options. We could implement changes on our own, since it's open source, but we can also request potential updates from Archivematica or provide them with funding to implement bigger changes, and

everyone benefits. A good example of this is Archivematica's integrations with other systems. We use a variety of different repositories and access platforms in the Libraries (Islandora, DSpace, Dataverse, AtoM) that hold hundreds of thousands of files which require preservation, and these systems are used by many other institutions. As a result, there have been previous requests to develop integrations for all of these systems with Archivematica, and we benefit from these past developments as it makes for better, more unified workflows as all of our systems can "talk to" Archivematica and vice-versa.

9. *What are some of the limitations of Archivematica, and how do you handle records that cannot necessarily be preserved in a state/format that is functional enough to be accessible to users?*

It's not an "out-of-the-box" solution, though that can be a benefit or a limitation. The drawback is that to install it locally, it requires a specific set of skills that not everyone has, but the benefit is that it's more flexible than an out-of-the-box, vendor-based solution and there's the option to make changes when it doesn't quite fit your needs.

We typically appraise records through a method that considers different criteria, including preservability, which accounts for whether the format itself can be preserved with existing resources, including software/hardware, but also, if it can be preserved usefully (i.e., to what extent are the preservation and access outputs usable, understandable, accessible). If we can't preserve it in a useful, meaningful way with existing resources, then it's a question of whether we can preserve it by investing in other resources, and if so, would the outcome merit the resources? If the record has some value, but it's not unique, or it's not crucial that we keep it, but preserving it would be a very time-consuming, costly effort, then we most likely wouldn't

acquire/preserve it, and might recommend that the material be donated to another institution that might be interested and might be better equipped to support the preservation. If we determined that it was worth preserving, then we'd have to consider what functionality would be lost, and whether that impacts the value of the content. It always goes back to the same question: What are we seeking to preserve? Would the elements of functionality that are lost be considered a significant property, or could we still preserve a record that remains useful, usable and understandable over time?

Participant B

1. Why did the Digital Strategies department decide to utilize format migration as its principle method of preservation for legacy formats rather than alternatives such as emulation?

Our designated user communities desire informational content over representation. So, for instance if we migrate an old word processing document, we are interested in the data that object holds more than we are rendering it in such a way as to reflect its original/legacy display. That being said, I do realize the that we are prioritizing the needs of our contemporary user communities and not necessarily anticipating what future user communities might expect from digital archives. Keeping in mind that we also preserve the original object in it native format, we are thinking of future archivists who might want to perform actions on the object that we might not have envisioned.

2. Resources, such as the Internet Archive, provide access to emulators like DOSBox, licence-free early versions of Windows and Mac operating systems, and compatible legacy applications. Such tools have the potential to allow archives to provide users access to older digital records while preserving a greater degree of their original functionality. What are the concerns and limitations to utilizing such tools as a method of preservation?

I wonder how important the "functionality" of older digital records is for users. Presently, emulators are nice to have, but not a have-to-have to represent the information contained within the data. There's also an issue of using third-party online applications for some digital content. We also might consider that emulators of legacy Windows and Apple Operating Systems (OS) are either illegal (i.e. running an old OS on a new machine) or an emulator that pretends to behave/operate like a legitimate replica of the original OS or GUI. Let's be honest, emulation as

a preservation option is currently focused on the representation of an object as-it-was in the GUIs of old. Emulation is currently more aesthetic than it is evidential.

3. When handling collections of mixed digital media (flash drives, DVDs, CDs, 3.5 and 5.25-inch floppy disks) how does Digital Strategies determine priority regarding what is preserved? Would newer storage media be processed first or is emphasis placed upon older, and possibly more fragile media?

Research Services and Digital Strategies does not always directly set priorities for migration. Many of the decisions we make on this topic are done in collaboration with the University of Manitoba Archives & Special Collections. However, user demands and media fragility are two things we keep in mind when setting preservation priorities. In this case, we would ideally migrate data on floppy media first, however, optical media (in my opinion) is just as fragile.

4. Since Digital Strategies works with legacy media, such as floppy disks, what role does legacy hardware play in preservation and what are the challenges of working with obsolete equipment?

Legacy hardware is important. Typically, we think of legacy hardware in terms of peripherals (floppy drives, Zip drives, etc.) but I find that older PC hard drives often contain proprietary software that cannot easily be ported to a newer machine/operating system. In reality, archivists rely on creators to bring their data up to contemporary formats before transferring it to archives, but given that the traditional approach to acquisition takes place some years or decades after those records were originally created, there is often a huge technology gap to be bridged.

5. What factors were influential in the decision to acquire the Forensic Recovery of Evidence Device (FRED)?

The decision to acquire a FRED was pretty straightforward. Highlighting again the reality of archives acquiring records (especially from private individuals) at the end of a creator's career or even life, the need to constantly recover data from legacy platforms is inevitable. In this sense, safe, reliable and auditable data recovery should be common practice for contemporary archival institutions. That was the essence of my argument for acquiring a FRED.

6. While FRED can read legacy storage media and identify legacy file formats, migration is still the method of choice for long term preservation and access. What details, such as the original environment (hardware and software) required to interact with and manipulate the original legacy formats, are included in the archival description as part of the record's provenance? (For example, noting that a record created in WordPerfect 3.0 for MS-DOS was migrated to Word or PDF format).

Unfortunately, I am somewhat removed from the creation of archival description in my current duties. That being said, the migration logs generated by software such as Archivematica do record these types of actions as could/should be include within an archival description.

7. Interactive media often contain embedded links, animation, interactive components, or other specific requirements in order to function (the code in individual spreadsheet cells for example). What are the limitations concerning what content can be extracted and preserved and how do you account for any lost functionality afterwards? (Are users made aware of any lost functionality in the record they are viewing)?

Given the amount of code used to render data graphically, it is difficult to capture every detail of a record's functionality, as well as a faithful legacy representation, with any digital preservation approach. For example, Web crawls do not always capture CSS data and the migration of optical media containing video does not always support DVD menu/index options or other elements. It is my understanding that those archivists working with users explain such details.

8. Archivematica has a wide variety of tools for digital preservation and supports several standards. Why was Archivematica adopted by Digital Strategies and what are some of the most important features for the work you do?

Given that Artefactual Systems have developed their technology with archivists in mind, programs such as Archivematica are a logical solution for memory institutions. It has been my experience that Archivematica's user interface, action logs, tool customization and scalability are the main reasons for its wide-spread adoption by archives. Archivematica also has incredible import/export functionality that non-IT administrators can easily use.

9. What are some of the limitations of Archivematica, and how do you handle records that cannot necessarily be preserved in a state/format that is functional enough to be accessible to users?

The limitations of Archivematics are, and have been for some time, the chokepoints that occur when processing content through the system pipeline. As Archivematica uses open source tools, sometimes these tools just don't "talk to each other" like they should and sometimes results in a processing failure.

While I do understand the importance of preserving (migrating, emulating, etc.) digital records, it is more important to actually appraise, for archival value, the digital records archivists seek to preserve. In the recent past, archivists have focused far too much on preserving legacy records because of the object's fragility as opposed to its actual value. This "panic preservation" has resulted in many institutions preserving terabytes of data that, more often than not, contain no archival value beyond the age of the media or format.

Bibliography

Adshea, Anthony. “DisStorage 101: Replication vs Backup, and Synchronous vs Asynchronous.” Computer Weekly. Accessed June 19, 2020. .

Artefactual Systems Inc. “Descriptive Standards.” Accessed April 12, 2020. https://www.accesstomemory.org/en/docs/2.5/user-manual/overview/descriptive-standards/.

Artefactual Systems Inc. “Export XML.” Accessed April 12, 2020. https://www.accesstomemory.org/en/docs/2.5/user-manual/import-export/export-xml/#export-xml.

Artefactual Systems Inc. “Identification Tools.” Accessed February 19, 2020. https://www.archivematica.org/en/docs/archivematica-1.10/user-manual/preservation/preservation-planning/#identification.

Artefactual Systems Inc. “What is Archivematica?” Accessed April 15, 2020. https://www.archivematica.org/en/docs/archivematica-1.11/getting-started/overview/intro/#intro.

Artefactual Systems Inc. “What is AtoM?” Accessed April 12, 2020. https://www.accesstomemory.org/en/docs/2.5/user-manual/overview/intro/.

Aspray, William and Donald deB. Beaver. “Marketing the Monster: Advertising Computer Technology.” *Annals of the History of Computing* 8 no. 1 (1986): 127-43.

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

AVP. “About.” Fixity. Accessed March 27, 2020. https://www.weareavp.com/products/fixity/.

Bak, Greg. “How Soon Is Now? Writings on Digital Archiving in Canada From the 1980s to 2011.” *The American Archivist* 79 no. 2 (2016): 283-319.

Bak, Greg. “Media and the Messengers: Writings on Digital Archiving in Canada from the 1960s to the 1980s.” *Archivaria* 82 (2016): 55-81.

Bak, Greg. “Not Meta Just Data: Redefining Content and Metadata in Archival Theory and Practice.” *Journal of Archival Organization* 13 no. 1-2 (2018): 2-18.

Bak, Greg. “Trusted by Whom? TDRs, Standards Culture and the Nature of Trust.” *Archival Science* 16 no. 4 (2016): 373-402.

Baldwin, Betsey. “Confronting Computers: Debates about Computers at the Public Archives of Canada during the 1960s.” *Archivaria* 62 (Fall 2006): 159-178.

Baldwin, Betsey. “Stepping off the Paper Trail? Rethinking the Mainframe Era at the Public Archives of Canada.” PhD diss., University of Ottawa, 2006.

Bass, Jordan. “Getting Personal: Confronting the Challenges of Archiving Personal Records in the Digital Age.” Master’s Thesis, University of Manitoba/University of Winnipeg, 2012. MSpace http://hdl.handle.net/1993/5206.

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

Becker, Christoph. “Metaphors We Work By - Reframing Digital Objects, Significant Properties, and the Design of Digital Preservation Systems.” *Archivaria* 85 (2018): 6-36.

Belanger, Sylvain. “LAC Digital Optimization of Canada’s Collections.” Presentation. University of Manitoba. December 2, 2019.

Bodrato, Stefano, Fabrizio Caruso, Giovanni A. Cignoni. “Discovering Eastern European PCs by Hacking Them. Today.” *arXiv* (2019). https://arxiv.org/abs/1901.06922.

Bredenberg, Karin, Angela Di lorio, Eld Zierau and Bertrand Caron. “Understanding and Implementing PREMIS – A Tutorial at iPRES 2019.” PowerPoint Presentation. 2019 PREMIS Implementation Fair, Amsterdam. September 16, 2019. https://www.loc.gov/standards/premis/pif/2019/iPres-PREMIS-Tutorial-2019-Final.pdf.

Brown, Adrian. *Practical Digital Preservation: A How-To Guide for Organizations of Any Size*. London: Facet Publishing, 2014.

Canada. Bureau of Canadian Archivists. *Rules for Archival Description,* Revised Version – 2008. 2008.

Canada. Privy Council office. “Report 4: Paperwork and Systems Management.” In *The Royal Commission on Government Organization. Vol. 1: Management of the Public Service / J. Grant Glassco, Chairman*. Ottawa: Queen’s Printer, 1962. http://publications.gc.ca/collections/collection_2014/bcp-pco/Z1-1960-4-1-5-eng.pdf.

Caplan, Priscilla. “The Florida Digital Archive and DAITSS: A Working Preservation Repository Based on Format Migration.” *International Journal on Digital Libraries* 6 no. 4 (2007): 305-11.

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

Chen, Su-Shing. “The Paradox of Digital Preservation.” *Computer* 34 no. 3 (March 2001): 24-8.

The Computer Chronicles. “The Computer Chronicles - Secrets of Word Perfect (1992).” YouTube video. Posted by “The Computer Chronicles.” August 1, 2013. https://www.youtube.com/watch?v=kHel_aqMOb8.

Computer History Museum. “Microsoft Software Catalog.” Accessed April 08, 2020. https://www.computerhistory.org/collections/catalog/102665404.

Connell, Matthew. “Computers as Historical Artefacts.” *Australian Journal of Multi-Disciplinary Engineering* 4 no. 1 (2006): 39-44.

Consultative Committee for Space Data Systems (CCSDS). *Reference Model for an Open Archival Information System (OAIS)*. 2012.

Cook, Terry “Electronic Records, Paper Minds: The Revolution in Information Management and Archives in the Post-Custodial and Post-Modernist Era.” *Archives and Manuscripts* 22, no. 2 (November 1994): 300–29. Reprinted in *Archives and Social Studies: A Journal of Interdisciplinary Research* (March 2007): 405–49.

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

Cook, Terry. “Paper Trails: A Study in Northern Records and Northern Administration, 1898-1958.” in *For Purposes of Dominion: Essays in Honour of Morris Zaslow*, edited by Kenneth S. Coates and William R. Morrison, 13-35 York: Captus University Publications, 1989.

Council of Prairie and Pacific University Libraries (COPPUL). “WestVault.” Accessed April 24, 2020. https://coppul.ca/westvault.

Craig, Barbara L. Craig. “Machines, Methods, and Modernity in the British Civil Service, c. 1870-c.1950.” *Journal of the Society of Archivists* 30 no. 1 (2011): 63-78.

Dancy, Richard. “RAD Past, Present, and Future.” *Archivaria* 74 (2012): 7-41.

Dappert, Angela, Sébastien Peyrard, Carol H. Chou, and Janet Delve. "Describing and Preserving Digital Object Environments." *New Review Of Information Networking* 18 no.2 (November 2013): 106-73.

Dietrich, Dianne, Julia Kim, Morgan McKeehan, and Alison Rhonemus. "How to Party Like it's 1999: Emulation for Everyone." *code[4]lib Journal* 32 (2016). https://journal.code4lib.org/articles/11386.

Digital Intelligence. "Overview." Training. Accessed April 20, 2020. https://digitalintelligence.com/training/overview.

"DOS Tutorial Commands for Beginners." YouTube video. Posted by "PhilsComputerLab." October 7, 2014. https://www.youtube.com/watch?v=MWiOeV-Fh7U.

Eastwood, Terry. "Appraising Digital Records for Long-Term Preservation." *Data Science Journal* 3 no. 20 (December 2004): 202-8.

Engle, Erin. "Authenticity Amidst Change: The Preservation and Access Framework for Digital Art Objects." *The Signal*, Library of Congress, Posted December 21, 2015. https://blogs.loc.gov/thesignal/2015/12/authenticity-amidst-change-the-preservation-and-access-framework-for-digital-art-objects/.

File Information Tool Set (FITS). "FAQ." Accessed April, 22, 2020. https://projects.iq.harvard.edu/fits/fitsfaq.

Fisher, Tim. "What is an ISO File?" *Lifewire*. Last Updated April 13, 2020. https://www.lifewire.com/iso-file-2625923.

Floridi, Luciano. *Information: A Very Short Introduction.* Oxford: Oxford University Press, 2010.

The Frank-Ratchye STUDIO for Creative Inquiry. "Warhol Amiga Report." (March 2014) Accessed October 26, 2019. http://studioforcreativeinquiry.org/public/warhol_amiga_report_v10.pdf.

Franks, Pat, and Nancy Kunde. “Why METADATA Matters.” *The Information Management Journal* Vol. 40 No. 5 (2006): 55-61.

Frisz, Chris, Geoffrey Brown and Samuel Waggoner. “Assessing Migration Risk for Scientific Data Formats.” *The International Journal of Digital Curation* 7 no. 1 (2012): 27-38.

Galloway, Patricia. “Personal Computers, Microhistory, and Shared Authority: Documenting the Inventor–Early Adopter Dialectic.” *IEEE Annals of the History of Computing* 33 no. 2 (2011): 60-74.

Galloway, Patricia. “Retrocomputing, Archival Research, and Digital Heritage: A Computer Museum and iSchool Collaboration.” *Library Trends* 59 no. 4 (Spring 2011): 623-36.

Gavrel, Sue. “Preserving Machine-Readable Archival Records: A Reply to John Mallinson.” *Archivaria* 22 (1986): 153-155.

Gilliland, Anne J. “Setting the Stage.” In *Introduction to Metadata*, edited by Murtha Baca. 3rd ed. Los Angeles: Getty Publications, 2016. http://www.getty.edu/publications/intrometadata/setting-the-stage/.

Grundig Business Systems. “Our History.” Accessed July 30, 2019. https://www.grundig-gbs.com/index.php?id=483&L=1.

Guttenbrunner, Mark, Christoph Becker and Andreas Rauber. “Keeping the Game Alive: Evaluating Strategies for the Preservation of Console Video Games.” *The International Journal of Digital Curation* 5 no. 1 (2010): 64-90.

Hackett, Yvette. “Preserving Digital History: Costs and Consequences.” In *Better off Forgetting?: Essays on Archives, Public Policy, and Collective Memory,* edited by Mona Homlund and Cheryl Avery, 124-139. Toronto: University of Toronto press, 2010.

Haigh, Thomas. “The Chromium-Plated Tabulator: Institutionalizing an Electronic Revolution, 1954–1958.” *IEEE Annals of the History of Computing* (2001): 75-104.

Inscho, Jeffery. “Hacking Vintage Technology to Simulate Time Travel.” The Andy Warhol Museum, Accessed October 26, 2019. https://www.warhol.org/hacking-vintage-technology-to-simulate-time-travel/.

Internet Archive. “About the Internet Archive.” Accessed July 12, 2019. https://archive.org/about/.

Internet Archive. “The ARCHIVE.ORG Manual Library.” Accessed February 23, 2020. https://archive.org/details/manuals.

Internet Archive. “Software Library.” Accessed May 11, 2020. https://archive.org/details/softwarelibrary

Internet Archive. “The Software Library: MS-DOS.” Accessed July 13, 2019. https://archive.org/details/softwarelibrary_msdos.

Internet Archive. “WordPerfect 6 for Windows.” Accessed February 23, 2020. https://archive.org/details/wordperfect6forw00crum.

International Council of Archives. *ISAD(G): General International Standard Archival Description* Second Edition. 2000.

KryoFlux. “Features.” Accessed October 23, 2018. https://www.kryoflux.com/?page=kf_features.

KryoFlux. “Technical Overview.” Accessed October 23, 2018. https://www.kryoflux.com/?page=kf_tech.

Lavoie, Brian. *The Open Archival Information System (OAIS) Reference Model: Introductory Guide, (2nd Edition)*. Digital Preservation Coalition, DPC Technology Watch Report 14-02 (October 2014). Accessed April 11, 2020. https://www.dpconline.org/docs/technology-watch-reports/1359-dpctw14-02/file.

Lazinger, Susan S. *Digital Preservation and Metadata*. Englewood, 2001.

Lee, Christopher A. "Open Archival Information System (OAIS) Reference Model." *Encyclopedia of Library and Information Sciences, Third Edition* (2010): 4020-30.

Lendino, Jamie. "Microsoft Windows Turns 30: A Brief Retrospective." *ExtremeTech*. Posted November 20, 2015. https://www.extremetech.com/computing/218336-microsoft-windows-turns-30-a-brief-retrospective.

Levi, Charles. "Five Hundred 5.25-Inch Discs and One (Finicky) Machine: A Report on a Legacy E-Records Pilot Project at the Archives of Ontario." *Archivaria* 72 (2011): 239-246.

Liebetraut, Thomas, Klaus Rechert, Isgandar Valizada, Konrad Meier and Dirk von Suchodoletz. "Emulation-as-a-Service – The Past in the Cloud." *2014 IEEE 7th International Conference on Cloud Computing*, Anchorage, AK, (2014): 906-913. 10.1109/CLOUD.2014.124.

Lunar, Steven. ""Do Not Fold, Spindle or Mutilate": A Cultural History of the Punch Card." *Journal of American Culture* 15 (1992): 43-55.

MacNeil, Heather. "Archivalterity: Rethinking Original Order." *Archivaria* 66 (Fall 2009): 1-24.

Mahoney, Michael S. "The Histories of Computing(s)." *Interdisciplinary Science Reviews* 30 no. 2 (2005): 119-135.

Mallinson, John C. “Preserving Machine-Readable Archival Records for the Millenia.” *Archivaria* 22 (1986): 147-152.

McKemmish, Sue, Glenda Acland, Nigel Ward and Barbara Reed. “Describing Records in Context in the Continuum: The Australian Recordkeeping Metadata Schema*.” *Archivaria* 48 (1999): 3-37.

McLellan, Evelyn Peters. “Selecting Formats for Digital Preservation: Lessons Learned during the Archivematica Project.” *Information Standards Quarterly* 22 no. 2 (2010): 30-3.

National Archives, The. “DROID Signature Files (Archive).” Accessed February 11, 2020. https://www.nationalarchives.gov.uk/aboutapps/pronom/droid-signature-files.htm.

National Archives, The. “PRONOM.” Accessed February 11, 2020. http://www.nationalarchives.gov.uk/PRONOM/Default.aspx#.

The National Archives. “Information Resources.” Accessed February 11, 2020. https://www.nationalarchives.gov.uk/aboutapps/PRONOM/tools.htm.

Naugler, Harold. *The Archival Appraisal of Machine-Readable Records: A RAMP Study with Guidelines.* Paris: United Nations Educational, Scientific and Cultural Organization, 1984.

Nesmith, Tom. “Still Fuzzy, But More Accurate: Some Thoughts on the “Ghosts” of Archival Theory,” *Archivaria* 47 (1999): 136-150.

Niu, Jinfang. “Original Order in the Digital World.” *Archives and Manuscripts* 43 no. 1 (2015): 61-72.

“Dr. William O. Pruitt, Jr.” *Winnipeg Free Press Passages*. Accessed July 28, 2019. https://passages.winnipegfreepress.com/passage-details/id-158565/PRUITT_WILLIAM.

oldcomputerbooks. “Results: Apple.” Accessed April 07, 2020. https://www.oldcomputerbooks.com/advSearchResults.php?action=search&orderBy=relevance&category_id=0&keywordsField=Apple.

O’Neil Adams, Margaret. “Punch Card Records: Precursors of Electronic Records.” *The American Archivist* 58 no. 2 Special Issue on Case Studies (Spring, 1995): 182-201.

PREMIS Editorial Committee. *PREMIS Data Dictionary, Version 3.0.* June 2015.

Preservica. “About Us.” Accessed February 11, 2020. https://preservica.com/about.

PBU (Power Back-Ups) Products. “Getting To Know: WordPerfect 6.0 for DOS - VHS (1993).” YouTube video. Posted by “member berries.” August 10, 2015. https://www.youtube.com/watch?v=jCVGRePLrCA.

Rosenthal, David S. H. “Bit Preservation: A Problem Solved?” *The International Journal of Digital Curation* 1 no. 5 (2010): 134-148.

Rosenthal, David S. H. “Emulation & Virtualization as Preservation Strategies.” LOCKSS Program, Stanford University Libraries (2015). https://digital.library.unt.edu/ark:/67531/metadc799755/m2/1/high_res_d/rosenthal-emulation-2015.pdf.

Rothenberg, Jeff. “Ensuring the Longevity of Digital Documents.” *Scientific American* 272 no. 1 (1995): 42-7.

Roy, Jordan. “Preserving Interactivity: Towards Next Generation Digital Preservation Philosophy and Systems.” Master’s Thesis, University of Manitoba/University of Winnipeg, 2019. MSpace (http://hdl.handle.net/1993/34406).

Schlieder, Christoph. “Digital Heritage: Semantic Challenges of Long-term Preservation.” *Semantic Web* 1 no.1 (2010): 143-7.

Scott, Jason. “Early Macintosh Emulation Comes to the Archive.” Internet Archive Blogs. Posted April 16, 2017. https://blog.archive.org/2017/04/16/early-macintosh-emulation-comes-to-the-archive/.

Scott, Jason “Internet Archive Does Windows: Hundreds of Windows 3.1 Programs Join the Collection.” Internet Archive Blogs. Posted February 11, 2016. https://blog.archive.org/2016/02/11/internet-archive-does-windows-hundreds-of-windows-3-1-programs-join-the-collection/.

Shepherd, Elizabeth and Charlotte Smith. “The Application of ISAD(G) to the Description of Archival Datasets.” *Journal of the Society of Archivists* 21 no. 1 (2000): 55-86.

Sprout, Bronwen and Mark Jordan, “Archivematica As a Service: COPPUL’s Shared Digital Preservation Platform.” *Canadian Journal of Information and Library Science* 39 no. 2 (2015): 235-44.

Stachniak, Zbigniew. “Red Clones: The Soviet Computer Hobby Movement of the 1980s.” *IEEE Annals of the History of Computing* 37 no. 1 (2015): 12-23.

Suchodoletz, Dirk von, Klaus Rechert and Isgandar Valizada. “Towards Emulation-as-a-Service: Cloud Services for Versatile Digital Object Access.” *The International Journal of Digital Curation* 8 no. 1 (2013): 131-142.

Suominen, Jaakko and Jussi Parikka, “Sublimated Attractions – The Introduction of Early Computers in Finland in the Late 1950s as a Mediated Experience.” *Media History* 16 no. 3 (2010): 319-340.

Takhteyev, Yuri and Quinn DuPont "Retrocomputing As Preservation and Remix." *Library Hi Tech* 31 no. 2 (2013): 355-70.

Taylor, Hugh. "Transformation in the Archives: Technological Adjustment or Paradigm Shift?." *Archivaria* 25 (1987-88): 12-28.

TechCrunch. "George R.R. Martin Writes Everything In WordStar 4.0 On A DOS Machine." Accessed February 13, 2020. https://techcrunch.com/2014/05/14/george-r-r-martin-writes-everything-in-wordstar-4-0-on-a-dos-machine/.

Trace, Ciaran B. "Beyond the Magic to the Mechanism: Computers, Materiality, and What It Means for Records to Be "Born Digital"." *Archivaria* 72 (2011): 5-27.

University of Manitoba Libraries and Special Collections. "William O. Pruitt Jr. fonds." Finding Aid. Accessed July 28, 2019. https://main.lib.umanitoba.ca/william-o-pruitt-fonds.

University of Manitoba Archives and Special Collections. "William O. Pruitt: An Inventory of His Papers at the University of Manitoba Archives & Special Collections." Last Modified April 22, 2014. http://www.umanitoba.ca/libraries/units/archives/collections/complete_holdings/ead/html/pruitt_10_112.shtml#series10.

Van der Hoeven, Jeffery, Bram Loham and Remco Verdegem, "Emulation for Digital Preservation in Practice: The Results." *The International Journal of Digital Curation* 2. No. 2 (2007): 123-32.

Van der Hoeven, Jeffery. "The Need for Emulation Services." *De Gruyter* 35 No.4 (2012): 235-9.

Vielfaure, Natalie. “Digital Preservation: How UM Libraries are Confronting the Challenges of Record Keeping in a Digital World.” *RDRS Blog.* Posted November 29, 2018. https://libguides.lib.umanitoba.ca/researchservices/rsdsblog/Digital-Preservation-How-the-Libraries-are-Confronting-the-Challenges-of-R.

Vielfaure, Natalie. “Have you Heard About FRED?.” *RDRS Blog.* Posted November 15, 2018. https://libguides.lib.umanitoba.ca/researchservices/rsdsblog/Have-you-Heard-About-FRED.

Walch, Victoria Irons. Compiler. “Index.” In *Standards for Archival Description: A Handbook*. The Society of American Archivists. 1994. Accessed April 10, 2020. http://www.archivists.org/catalog/stds99/index.html.

Wandt, Adam Scott. “Digital Forensics: Hardware.” May 12, 2013. https://www.youtube.com/watch?v=-qPVtaxJWv4&feature=youtu.be.

Whalen, James M. ““Out, Damned Spot!” The Staining of the Proclamation of the Constitution Act, 1982.” *Archivaria* 61 (2006): 289-298.

WikiHow. “How to Use MS DOS.” Accessed August 8, 2019. https://www.wikihow.com/Use-MS-DOS.

William O. Pruitt Jr. “Electronic Records.” 1988-1996. EL 25 PC 28, Box 1-2. William O. Pruitt Jr. fonds. University of Manitoba Archives and Special Collections, Winnipeg, Manitoba, Canada.

Woodley, Mary S. “Metadata Matters: Connecting People and Information.” Revised by Murtha Baca. In *Introduction to Metadata*, edited by Murtha Baca. 3rd ed. Los Angeles: Getty Publications, 2016. http://www.getty.edu/publications/intrometadata/metadata-matters/.

Woods, Kam and Geoffrey Brown, "Assisted Emulation for Legacy Executables." *The International Journal of Digital Curation* 5 no. 1 (2010): 160-171.

Yeo, Geoffrey Yeo. "Bringing Things Together: Aggregate Records in a Digital Age." *Archivaria* 74 (2012): 43-91.