Preserving Interactivity:

Towards Next Generation Digital Preservation Philosophy and Systems

by

Jordan Grant Kaczor Roy

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

MASTER OF ARTS

Department of History (Archival Studies)

Joint Masters Program

University of Manitoba/University of Winnipeg

Winnipeg, Manitoba

Table of Contents

List of Figures

Abstract

New emulation technologies have emerged in the form of Emulation as a Service, and are currently in the late stages of development. These projects have the potential to have positive implications for digital preservation and access. The nature and scope of these projects however, requires examination. A look at what these technologies can do, as well as how they fit into current archival systems, is a necessity. As these technologies allow for the presentation and maintenance of previously neglected properties of the record, there emerges a need to understand why these interactive aspects are important. Archivists will need to begin to think about how these technologies function and can fit into current and prospective archival methods, in order to most effectively make use of them.

Acknowledgements

I would like to acknowledge Professor Tom Nesmith who introduced me to the archival world years ago. Your honest passion for the archives served to inspire more than my own interest. Since those first steps I have made many of my own archival stories that will stay with me for a lifetime. I would also like to thank my advisor Professor Greg Bak, who has helped me write and edit this thesis, as well as all the other papers that I wrote during my classes in attempts to form a idea. You shared with me a curiosity in all aspects of our digital world that made this work possible.

I would like to thank those that I have worked with in the various memory institutions that allowed me to explore and make mistakes. You all showed me that there is more to the archives than processing records. There are too many to name, but working alongside each of you showed me the value of archival work.

I would like to thank my family for your support, you were there for me with a warm meal and encouragement when I had to tighten my belt, even if I was so focused on an idea it was all I could talk about till I figured it out.

A special thank you to Kris Driver your benevolent criticism and unrelenting curiosity gave me an example I will cherish and seek to emulate forever.

Finally, to my wife Megan who I forced to read this at least twice.You were always there to listen and challenge me. Your perspective and knowledge have widened my own. You truly are my other half.

Introduction

Video games are meant to be played, not just looked at like a painting, watched like a movie, or listened to like music. It is possible to go online and watch someone play a game, but nothing quite replaces the joy of figuring out a complex puzzle, the thrill of getting to the safe room after overcoming waves of zombies, or the frustration when a natural disaster threatens the digital citizens of one's meticulously built city. People play games because of the way the interaction moves them. They are charging into battle, and they are building that house with their actions. Without that input, much of the game is lost.

Most digital records that an archivist will encounter will not be as entertaining, or have such an abundance of novelty. Most will be everyday records such as text documents, spreadsheets, scripts, research data, or emails. However, the way they are designed to be manipulated and altered is just as important. We use a mouse to click and make alterations to a spreadsheet, then use a keyboard to input data.

These interactions inform the decisions we make when using the software. Using another example from games, having a joystick helps a person more expertly control a flight simulator. A person flying a virtual plane with a joystick would make different decisions than they would have had they been forced to use the directional keys on their keyboard. An actual example can be found in the case of the Chaste 3.1 software, which is a tool used to model "lung structure, dynamics and functions."[1] The proper functioning and interfacing of this software is essential, as it has been used and citied in multiple scientific studies. The data, which any analysis would be based on, must be reproducible in order to verify the reliability of the results of an analysis. As Chaste 3.1 is medical science, proper interactive use of any data, could aid in reproducibility of

[1] "Lung Chaste: developing software for multiscale lung ventilation simulations," University of Oxford Department of Computer Science, accessed July 3, 2019, https://www.cs.ox.ac.uk/chaste/lung_index.html.

the results. This could have positive consequences for people dealing with lung problems or disease. A person's understanding of a game, of scientific data, or even a spreadsheet file is diminished by their inability to properly interface with those digital records. Conclusions made from interactions with a record through an incorrect, insufficient interface would be lacking and likely be open to critique.

Experience leads to understanding. Like the flight simulator, having interactive access to any digital record allows the user to see why decisions were made, and why the record was written, collected, drawn, or arranged the way it was. Archival scholar Tom Nesmith wrote that "our understanding of reality is powerfully shaped by the particular forms and media of communications in which we are immersed."[2] Without this immersion, we are forced to fill in what gaps are left by the lack of interactivity, and are forced to use our imaginations with their biases and limitations, which could in turn lead to misunderstandings, having unintended, yet very real, consequences for people who have an interest or could be impacted by those conclusions.

The value of user experience goes beyond the realm of archives. A person learning about musical instruments should make an attempt to play the ones they are studying. They should carry it around, take care of it, and perform the tuning and maintenance it may need. They would learn from the weight of it, the way it forces their fingers to move, how it forces them to stand or sit, and how it changes their actions. The experience will have taught them much, even if they play badly. However, researchers cannot do this with all instruments. There are some that we know of, but no longer have examples of or plans on how to make one. There were no examples of a Gue, a bowed lyre from the Shetland Islands. Any writing relating to them would be

[2] Tom Nesmith, "Seeing Archives: Postmodernism and the Changing Intellectual Place of Archives," *American Archivist*, Vol. 65 (Spring/Summer 2002), 29.

speculative. That is until someone reconstructed one in 2009. Many others are lost forever, such as the Norwegian bowed lyre, the Giga.[3]

This value of experience is dependant on the existence of those objects. That is to say that, more than merely making the digital records available interactively, an example of the record still must exist for that to happen. As this is true of musical instruments, it is true of digital records. This is not a problem that can be solved by a new technology, it is simply making sure digital objects are put through proper preservation before entropy takes its toll on yet another hard drive.[4] Fortunately, there are still examples of much computing technology and much of our digital heritage is still in existence. There is an opportunity at the present time to maintain the interactive elements of digital records.

It is for these reasons that this thesis argues that the interactive elements of a digital record are an important part of that record. Archivists should strive to make a record available with all its interactivity present, while maintaining the integrity of the record. To do this, archivists working with digital materials should be aware of the most current technologies and methods for preserving and accessing these records, as well as how they fit into archival workflows.

The first chapter of this thesis serves to lay out the historical and philosophical foundation to my work. In this chapter, I will explore two aspects of digital interactivity. The first part of this chapter, will argue for the value of that interactivity. I frame my argument using the work of Marshal McLuhan. Rather than see a record's medium as merely a means to an end,

---

[3] These are not necessarily relevant beyond serving to make an argument. Otto Emanuel Andersson, *The Bowed-Harp: A Study in the History of Early Musical Instruments* (New York: AMS Press, 1930), 165-169; "Bowed Lyre ~ Shetland Gue," Kate & Corwen Ancient Music Ancient Instruments, accessed, July 3, 2019, http://www.ancientmusic.co.uk/string.html#gue.

[4] It is not the explicit purpose of this thesis to make a case for the haste of digital preservation initiatives, rather to make cases for a more nuanced approaches that maintain the interactivity of those digital records. Though the author is of the opinion that an expansion digital preservation initiatives would be beneficial.

McLuhan and archival scholar Hugh Taylor argue that the medium is part of that record's message, and not merely a disposable means to an end.

The second part of this chapter will look at how our digital media's interactive aspects are manifested in history. I assume that a digital record's interactive elements are not subtle abstract concepts, but physical manifestations of interaction with our machines. What makes a record interactive can be answered simply by looking at how one would use that record. However, there is an immense variety in how computers can be interacted with. I will limit my analysis to a number of popular forms that will illustrate my point: the mouse and keyboard, graphical interfaces, hypertext, and web browsers. These seem obvious; but it is for that reason that they are the most necessary to examine. Preservationist Ala Rekrut noted that "Overwhelmed by the sheer volume of the records we create, manage, and consult, we risk underestimating the cultural choices that they manifest."[5] We can easily miss the consequences of a thing as common as a mouse. I argue in this section that an archivist should understand not just the content of a digital object, but have an understanding of the processes that exist as parts of that record, that a person in most other positions would properly take for granted.

The second chapter moves on from the first and asks: if the interactive elements of digital records are important to maintain and make available, how can this be done, and what does this mean for an archivist? To answer this, chapter two will be broken up into two parts. The first of these will look at the technical aspects of the question. For this I argue that emulation is the best tool that archivists have to maintain the interactivity of a digital record. It is not the only tool however, and it should be used alongside other tools - such as various forms of migration. I weigh the strengths and weaknesses of these technical options. I argue that not every digital

[5]Ala Rekrut, "Material literacy: reading records as material culture," *Archivaria* 60 (2006), 35.

record needs to be accessed in its interactive environment every time, as different records have multiple different properties, not all of which are necessary in every case. Thus maintaining interactivity becomes a matter of appraisal, the extent of which is largely a matter of the archivist's discretion.

The archivist would be emulating digital records within a larger archival structure that they and their colleagues use to better work with the collections in their archive. That is why the second part of chapter two will look at how emulation fits into the two dominant archival approaches which fall within the scope of this work: the life cycle and the continuum models. Though each are in use, to various extents, I will look at each to show how a digitally emulated record exists with these systems.

The final chapter will focus on the more practical concerns in implementing emulation. As such, the purpose of this chapter is to bring awareness of the most recent technology and methods to an archival audience. Many archivists do not have the time to examine every aspect of a technology or become experts in web design, XML, or writing software. Accordingly this thesis will seek to give each of the approaches a fair analysis so as to let archivists decide for themselves.

The first half of this chapter will look at successful implementations of emulation as an access tool. The first of these are Salman Rushdie's digital records, which are available at the Manuscript, Archives, & Rare Book Library at Emory University. This project, which was led by Laura Carrol, represents an ad hoc approach with limited scope. The other major example in the first half of the chapter is that of the Internet Archive's Software Library, which represents the application of open source software on a large scale. I will suggest that both of these case studies, while successful, are examples of methods that are not practicable as a standard archival tool.

This thesis takes a tentatively optimistic tone however, as there are a number of technologies in the late stages of development that have the potential to provide scalable and reproducible emulation of digital objects. The Emulation and Software Preservation Infrastructure (*EaaSI*) being run out of Yale, and the Olive Executable Archive being built at Carnegie Melon University in collaborative with IBM Research, show a lot of promise and have working examples of their software. These projects seek to provide emulation as a third party service with the expressed goal of providing archives, and other memory institutions, with the capability to offer interactive access to digital records. I will examine the technical aspects of both of these systems and examine the implications that their approaches would have on archival work.

It is my aim to nest these technologies in the context of archival theory. I am of the opinion that theory and practical application must be put side by side and uttered in the same breath. It is beneficial for one to have an understanding as to why a thing is being done. EaaSI and Olive do represent potential practical alternatives, but any decision made with regard to digital preservation and access, aside from a do-nothing approach, will require expenditure.[6] It is for this reason that application and implementation of the technologies that will be discussed in chapter three rely on the archival context in chapter two, which relies on the philosophical foundations of the first chapter. I am not arguing that archivists should drop everything and start implementing emulation. This thesis serves as a way of thinking about the interactive elements of digital records, how they fit into archival workflows, as well as a few places to start considering the preservation of those elements.

[6] One could argue that a "do nothing" approach merely moves the expenditures to future archivists.

To that end this thesis is aimed both at those archivists that may not have a lot or any experience with the digital, as well as those with experience. As I look at some of the newer approaches it is my hope that both the computer illiterate and tech savvy in the archival world find interest and explore these topics and approaches. I encourage the reader to follow the links provided in the footnotes. Many of the videos, such as Douglas Engelbart's Mother Of All Demos, wonderfully illustrate how modern interactive systems took shape. Many of the links in the third chapter are valuable as they contain interactive examples of the software that I am describing. I take time to describe in detail these interactive processes, but as one of the underlying themes of this thesis is that "seeing (or rather interacting) is believing" it is best to play with these tools, examples, and demos oneself. This engagement will further serve to demystify these processes.

Many are afraid of the digital; it is complex and not instantly evident how or why it works. Thus, to many it becomes, or seems like, something that others take part in. I will at times be using very technical language, but it is necessary to go into the technical aspects of much of what I discuss in order to see the nuance of what certain technologies entail. It does matter how a record is accessed. A record can be seen as part of a story, and it matters how a story is told. This thesis is not about what we are saying, but how to let the stories, actions, and ideas be told with the utmost of nuance. I hope to convince the reader of the value of the interactive aspects of digital records, and to make digital interactivity seem like it is not just the purview of another department.

Chapter One:
Archivists should have thermostats: Narratives of Interactivity and Their Relevance to Archivists

It is easy to say that interactivity is a property that all records share. This statement belies the variety of media and ignores that for some records interactivity maintains a different space than others. It is one of the goals of this chapter to look at and define the nature of the interactivity of digital records, and in doing so show the value of maintaining a digital record's interactivity for access. This is essential, as the following two chapters will assume that maintaining interactivity is important. Its importance is not entirely obvious: thus it is necessary to discuss what is meant for a record to have interactivity as a relevant property. This definition should be able to stand as a tool for factoring interactivity as worthy of being maintained during appraisal. To do this I will rely on the work of Marshal McLuhan and Hugh Taylor whose work emphasised the examination of media.

Knowing that digital records can be interactive does not tell one why they are interactive, or why certain types of interactions are important. It is for this reason that I will examine the history of the most ubiquitous technologies and concepts that are found in interactive digital systems of today. This chapter will not merely be a history of computing. It will focus on the systems themselves, how they were constructed, and how they made interactivity core to user experience. Ultimately, this chapter will show how people interact with their machines and how we as a society constructed, and still construct, those modes of interaction. To the archivist this will suggest that it is necessary to save not just the what (the things and content) but the how (the interactivity) of these environments.

Part One: What is Meant by Interactivity

It can be argued that all media is interactive in that we interact with it in some way. We read a book, look at a picture, or click on a video online. We are interacting with them when we press the play button on a video or our remote control, or turn the page of a book. We then engage with the content by reading or watching as we lay on our couch, or watch with friends in a theatre. Much of the more recent media forms do not stop at this level of engagement. With media such as video games and web environments, the user has become a creator, or at least a co-creator: one is no longer the passive consumer of the media they engage with. The relationship between the user, the interface, and the content mirrors the estimation in Hugh Taylor's statement that, "the interplay between the medium and the receiver creates a communications environment over and above the content of the message and thereby becomes a message in itself."[7] That is to say that the way a person interacts with a record may be part of what can be learned from it. For archivists this means that the way in which people interact with records such as websites, games, programs, and desktop office suites can render relevant information about the record.

A man-computer interdependence has long been the norm in the lives of individuals in our society. As technology became more integrated into our lives, the amount of digital records that an archivist was responsible for kept pace accordingly. Archivists are not just seeing digital documents but executable scripts, networked environments, digital art, and can expect other programs and formats that we yet have not seen. Digital innovation continues and shows no end to its evolution. Whatever the mandate of the archive, the archivist has a responsibility to

[7]Hugh A. Taylor, "The Media of Record: Archives in the Wake of McLuhan" in *Imagining Archives: Essays and Reflections,* ed. Terry Cook, and Gordon Dodds (Lanham: Society of American Archivists and Association of Canadian Archivists in Association with Scarecrow Press, 2003), 64.

maintain a pace with the idiosyncrasies of the records that they should expect to encounter, lest the records arrive and they cannot be put through proper preservation.

Preservation aside, maintaining a record's interactivity is theoretically simple with more traditional records. One has simply to provide the original book or document on paper; even a photocopy will offer the same experience of reading from a page. However the interactive elements, that is to say the physical presence or method of conveying the data, are not always necessary for every consultation of the record. A book, for example need not be held, as its content can be accessed on a desktop, or an e-reader without substantially changing the meaning of the content. In many ways interactivity may not matter to the absorption of the content. I will forgo any debate around content vs. medium and say that it is circumstantial. I am also ignoring any debates around transcendence or romanticism that may be associated with an authentic article. [8]

What happens however when there is an element to the record for which the user's input affects the content of the record? A prime example of this would be a pop-up book, complete with moving parts. Though not a common item, a pop-up book when scanned would lose much of its functionality and render the record incomplete. The interactive parts of the book, the open and close doors, and the movable characters would be rendered immovable. Notes could be made in an attempt to signify which parts could move in a facsimile, but some of the content would be lost.

---

[8] Some research will require specific editions or copies, such as originals, to verify and cross-reference the various works. There is also much that can be learned from a material analysis or physical copies, such as more accurate dating, and confirming authenticity. Additionally archivists should allow people to have that experience with an original, or specific format so as to gain a greater understanding of that record, as long as there are no preservation concerns. Finally, individuals vary in their ability to absorb content, making different formats more effective for different people. For example someone with dyslexia may find transcribed audio more useful than a written copy. I am simply arguing on the grounds that a person's interactive access to a record conveys tangible knowledge that a researcher can use, rather than as a way to satisfy vague transcendent ends.

The same goes for certain digital objects. Like a pop-up book certain digital objects require, or at least are aided by, their interactive elements being present and functional. Like a book, a text file can be copied and its format changed and even printed and the intended function of the content, in this case the text to be read, does not change. Depending on the format, bit level data would be changed and, if printed, the metadata and bit level data would be lost, but it could still be read, although its context was erased.

However not all digital objects are documents. A program is a string of code, but to understand it, it must be run. A good example is that of the spreadsheet. A spreadsheet file is not intended to be a simple table. Programs like MSExcel and VisiCalc allow one to take a table of data and manipulate it. One can sort and filter data, while adding or removing it before arranging a selection of the data into graphs. To take a table of data and migrate it to a text file or have it printed off would remove the ability to engage with the data in these ways.

Thus to define this quality that a record may display, we can say that digital record's "interactivity", with relevance to an archival setting, are those aspects of a digital record where a user's input is integral for use and comprehension, and where a resulting output, culminating in a set of data to be consumed, is not the primary significance of the media. In other words, interactivity can be said to be an essential aspect of a record wherever the action of engaging with the media is deemed to be as important as the content of the media. I will explain what I mean by this in the next section.

So why is interactivity important? Why not settle for the content created as a result of our interactions with technology rather than the entire process itself? Emulation, though it remains the chief means of preserving digital interactivity, is expensive and complex. While there are very good arguments to disregard interactive elements in collections, I argue that those interactive elements that fall under the definition above are a necessary part of the record itself, constituting a record in itself.[9]

I take my reasoning from that of Canadian media studies scholar Marshal McLuhan, who saw the context of the content as being a statement of the culture in itself, and was interested in what media did to the people who used it.[10] To him people are "changed by the instruments that they employ." McLuhan's work sought the "principal of intelligibility" within each form of media that he had encountered in his time. That is to say he was looking to understand these forms of media and their effects, so as to bring them into an "orderly service."[11] McLuhan attempted to see how technologies such as the comic book, light bulb, and television affected the people in society; the changes that that they had on the whole society; and their consequences for the individual. To apply this mode of thought to an archival context would require an understanding of how the technologies, or the media, affects the reception of the message.

This is in reference to McLuhan's famous axiom that "the medium is the message". More clearly, he is saying that "The personal and social consequences of any medium…result from the new scale that is introduced into our affairs by each extension of ourselves."[12] Every new technology, be it a hammer, a book, or electrical light, will change the society that begins to utilize it. This is not a direct thing as there is no way to tell how a society will use a technology,

---

[9] There is another reason as to why but it can be summed up by three statements: it's cool, it's more interesting, and it's more fun. Though lacking the gravitas that is present in my arguments here, this point should not be readily disregarded. Holding people's attention is important. I will return to this point and develop it more fully in my conclusion.

[10] "Societies have always been shaped more by the nature of the media by which men communicate than by the content of the communication." Marshall McLuhan, Quentin Fiore, and Jerome Agel, *The Medium Is the Massage* (New York: Random House Inc.1967), 8. McLuhan talks further about the nature of his work at his 1974 lecture Living in an Acoustic World: Marshall McLuhan, "Marshall McLuhan 1974 – Full Lecture Living in an Acoustic World | University of South Florida," YouTube, 54:09, "mywebcowtube," May 2, 2016, https://www.youtube.com/watch?v=0l_ugK386QY&t=1465s.

[11] Marshall McLuhan, *Understanding Media: The Extensions of Man* (Cambridge: The MIT Press, 1994), 6.

[12] McLuhan, *Understanding Media*, 7.

and often a technology will be used for other than its desired effect. This can be seen in computer scientist Tim Berners-Lee's reaction to the use of the *WorldWideWeb* (WWW) he helped build, claiming "I was devastated," when asked about how humans have abused the technology.[13] Berners-Lee had little control over his invention's use. Humanity took it and built a giant network for study, and communication, as well as pornography and network attacks. As much as humanity changed the WWW, the WWW changed humanity.

Thus to McLuhan "it was not the machine, but what one did with the machine, that was its meaning in its message."[14] How people use a technology is just as important as the outcomes from the use of that technology. A book will force one to stop and read it, while having an electronic copy on a tablet would let a person transport it, copy it, search it more effectively. What is more, one can use a tablet to play an audio recording of the same text, allowing that person to wash the dishes, and walk to the supermarket while they are reading. This is an interactive distinction that is inherent to all technology, and all forms of records. This change in the medium has affected that person's day, and thus had changed the way they interact with society. The ability to have a computer in one hand allows everyone to consume that media, communicate, or play games in new ways. Our actions make up society, and as individuals we are limited in our action by what is possible. Emergent technologies allow us to act in modes that were previously not available, changing the needs, the wants and the interests of that society. McLuhan makes the claim that the "restructuring of human work and association was shaped by the technique of fragmentation that is the essence of machine technology."[15] This is not limited

---

[13]Katrina Brooker, ""I Was Devastated": Tim Berners-Lee, the Man Who Created the World Wide Web, Has Some Regrets," *Vanity Fair*, accessed July 3, 2019, https://www.vanityfair.com/news/2018/07/the-man-who-created-the-world-wide-web-has-some-regrets.
[14]Brooker, ""I Was Devastated.""
[15]McLuhan, *Understanding Media*, 8.

to what we would think of as technology: the medium can be any idea, resource, tool, or action anything that we interact with.

McLuhan asks what is the "content" of a medium, as opposed to its source? That which comes about as a result of the use of the medium being the "content". It is reasonable to say that the medium is the thing that produces the message; and one may assume that the message would be the "content". McLuhan would argue that the distinction between medium and "content" is a false one, as it is the "content" of the medium that in turn drives change and creates "involvement in depth."[16] This "content" becomes the medium, the thing that produces an effect. These effects of the medium are what drive people: "It is the medium that shapes and controls the scale and form of human association and action."[17] Thus the ends to which we strive are informed by the means by which we achieve them. This sounds lofty, but I am writing this on a computer screen which is informing the topic, and can tell one much about not only how I achieved this, but why. My actions are acting as the "content" of my tools. To again use the example, a look at a spreadsheet will tell us the numbers and figures, but interacting with and manipulating the spreadsheet will tell us what was needed to put that number together. Why were those pieces of data arranged in that way? The messages to and from the computer, the program, the business, and the job, came together to arrange that data in a specific way. The program allows the movement of the data in a way that leave the final static version of that data only one point in that record's history. The "content" of that file is the ability it granted that user to manipulate that data. Decisions may have been made based on how that data was arranged, or on the ability to view variable arrangements or outcomes based on calculations within the spreadsheet. This applies to the archival context as well. A finding aid containing records under

[16]McLuhan, *Understanding Media*, 9.
[17]McLuhan, *Understanding Media*, 9.

appraisal will be filtered and sorted, columns added and removed, then sorted again, all to look for the best way to make their appraisal decision.

I am not advocating for maintaining the interactivity of every digital record, although this could be done using emulation, which will be discussed as an archival tool in chapters two and three. As all records contain an interactive message, why not just emulate everything? This is a position taken by Jeff Rothenberg, though with more nuance, which will be discussed in the following chapter.[18] The most obvious argument against emulation is the cost and the value of completing such an endeavour. These criticisms would be well founded and correct to argue against such a plan. McLuhan's work gives a convenient distinction to decide what has value as an interactive object and what would serve society just as well as a static digital record. Archives will still have to decide which records are worth maintaining with their interactivity over the long term. Most digital objects will not require emulation, though some technologies are coming up that allow emulation to be done on a basis of need, as we shall see in chapter three. As one of the essential elements of any record is participation, the records where interactivity is deemed to be not necessary, will tend to be what McLuhan would describe as hot media. In relation to these, the records in most need of emulation are cool media.

McLuhan saw this distinction, between hot and cool media, as one of participation. In explaining this distinction McLuhan writes that "any hot medium allows of less participation than a cool one, as a lecture makes for less participation than a seminar, and a book for less than dialogue."[19] McLuhan defines participation as "completion by the audience," or how much

[18]Jeff Rothenberg, "Ensuring the Longevity of Digital Information," Rev ed. (1999): 1-18.
[19]McLuhan, *Understanding Media*, 23; when I was asked "why is hot "hot," and cool was "cool?"" One way I explained it to someone was that "cool" media is cool because you have to put energy into it. Hot media is prepared for you so you can just eat it now, but it may not fill you up.

involvement the viewer has to put forth to engage with the medium.[20] Like his comparison of the formulaic waltz and free flowing jazz dancing: “The waltz was hot, fast mechanical dance...In contrast the twist is a cool, involved and chatty form of improvised gesture.”[21] It is also a matter of the resolution of the medium, “the ear is given a meager amount of information. And speech is a cool medium of low definition because so little is given and so much has to be filled in by the listener.”[22] Hot and cool media can come down to the amount of input that a receiver or user has to provide in order to understand or create the message.

When we look at files on our computer we are engaging with a platform that is relatively cool. Most of the processes are happening off-screen, but the user is participating in the operating system and the programs that run in it. The user decides where to go and what to do, and their mind is filling in all the gaps as to what files mean, the structure of the file system, and the meaning of the various symbols. Few people will create an operating system, and therefore few archivists will see an example of an operating system as something worthy of keeping. What the individual does with a program is usually what is of importance: their writing, scripts, art, journals, and any creation that we might make. Like the appraisal of records on any other media an archivist has choices to make based on the format. Beta or VHS? Floppy or CD? If the archivist acquires a manuscript, they are committing themselves and the archive to handling, storing, and presenting it in a certain way. They could perform format migration to a digital or paper format on the manuscript, but the original may still have value for its authenticity and material information. This applies as well to digital files, as all digital formats are not equal, they should be handled according to their properties.

---

[20]McLuhan, *Understanding Media*, 23.
[21]McLuhan, *Understanding Media*, 27.
[22]McLuhan, *Understanding Media*, 23.

This is where McLuhan's work becomes important again. When we look at digital files with the same relative lens that we look at the digital and analog, we can still see a temperature spectrum.[23] A text file will can be participated with but the content can just be read. The medium is hot when compared to a record like a spreadsheet that invites you to sort and filter with the access copy. The records of an author of games may have beta examples of environments, technology and mechanics, that could be played in. When appraising for interactivity these cooler mediums have value. Gauging what level of participation a record might need upon access, would allow for researchers to better see the message of their medium.

Part Two: A Brief History of Interactivity

To further articulate what this means in regards to the distinction of interactive media, it is useful to look at the development of interactivity. Interactivity has its roots in the history of computing. A history of interactivity would constitute an aspect of that history, comprising a thread that winds its way through the evolution of those tools. These tools augmented human capabilities through advancements in calculation speed, and the performance of complex tasks. Ways of accessing and manipulating these processes and data needed to evolve in tandem. These devices and methods made through these processes and others, such as punch cards and microfilm, allowed the creation of machines of increasing complexity. As information technologist Ted Nelson claims, digital computing "could have come from anywhere, it just happened that the military needed a computing device for their ballistic tests and cryptography."[24] Tools like the UNIVAC, the first commercially available general purpose computer, and other early computers, did not yet signify a breakthrough in interactivity. People

[23] A metaphorical one.
[24] Max Allen, "Interview with Ted Nelson," Ideas CBC Radio, 1979, accessed July 5, 2019 https://archive.org/details/ideas-maxallen-tednelson.

had interacted with a machine, but the machine environment was limited often to its functions, much like a calculator, even an advanced one.[25] These tools did help make technological computation a more mature and established field, but the technology was still not democratic, as it required a great deal of expertise to use.

The records these early computers would have created would have existed on paper printouts, or on punch cards.[26] These outputs would have been able to be saved by standard archival methods, as most of these were in the form of paper. The records these machines would have created existed as the result of data being input, and then the processed data being output. This was done by a programmer working closely with the machine. This may entail preparing the machine for the input. For example, the ENIAC, one of the first digital stored program computers, had to be reprogrammed for every process it was to run: this involved rewiring much of the machine's hundreds of wires.[27] The interactivity was inherent to the calculation process. This may be as the computers did little else and were quite specialised for their tasks.

I am less interested in the characteristics of earlier computers than I am with the nature of the current digital environments that are already becoming part of the workload of even archives of the smallest scope. For even archivists with smaller collections have to ask how the things they have deemed archival will be interacted with and if that is worth saving. I will limit myself to the most basic of innovations, as there is no room to go into every aspect of digital interactivity. In reality, every aspect of the machine has an effect on the user experience. The RAM can make your movement in a system jerky or smooth depending on the amount you have available. The coding of the links between the processor and the various components has an

[25]Martin Campbell-Kelly, William Aspray, Nathan Ensmenger, and Jeffrey R. Yost, *Computer: A History of the Information Machine*, (Boulder: Westview Press, 2013), 98.
[26]Campbell-Kelly et al., *Computer,* 73-74.
[27]Campbell-Kelly et al., *Computer,* 70-72.

effect. There have been however, a few key technologies that have gone into creating the interactive environments that a majority of people have experience with today: the mouse, keyboard, screen interface, the development and design of graphical user interfaces (GUIs), and the hypertext systems that we move though every day. What is of interest is how and why these were envisioned, how the medium became the message.

Bush and the Memex:

A logical starting point can be found with Vannevar Bush, a computer scientist who in no small way influenced the subsequent developers in this paper. Bush, originally a professor at MIT in the early twentieth century, was known for his Differential Analyser, a device he built in 1931, which had the ability to solve numerous differential equations.[28] During the Second World War, Bush led at the Office of Scientific Research and Development, directing the US's scientific efforts.[29] His work directing the creativity and collaboration of large groups of individuals led him to see the value of a machine, not merely as a way to get a numerical result, rather as an "enlarged intimate supplement to his memory."[30] With the ability to store values and information, he saw that a machine that could act as not only a repository, but also as a way to edit, interact, and connect with the stored knowledge of others. With lessons learned in the course of his work of the power and benefit that science and technology brought to humanity, Bush published an article in *The Atlantic* titled "As We May Think."[31] This article was aimed at a post-war audience, whose creative energies he sought to move away from the destructive ends of war, to the anticipated questions that people focused on beneficial research will face.

---

[28]Campbell-Kelly et al., *Computer,* 49.
[29]Campbell-Kelly et al., *Computer,* 277.
[30] Vannevar Bush, "As We May Think," *The Atlantic,* (July 1945), accessed July 5, 2019, https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/.
[31]Bush, "As We May Think," Section 1.

Physicists he stated, with the creation of the atomic bomb, "have done their part on the devices that made it possible to turn back the enemy… Now, as peace approaches, one asks where they will find objectives worthy of their best."[32]

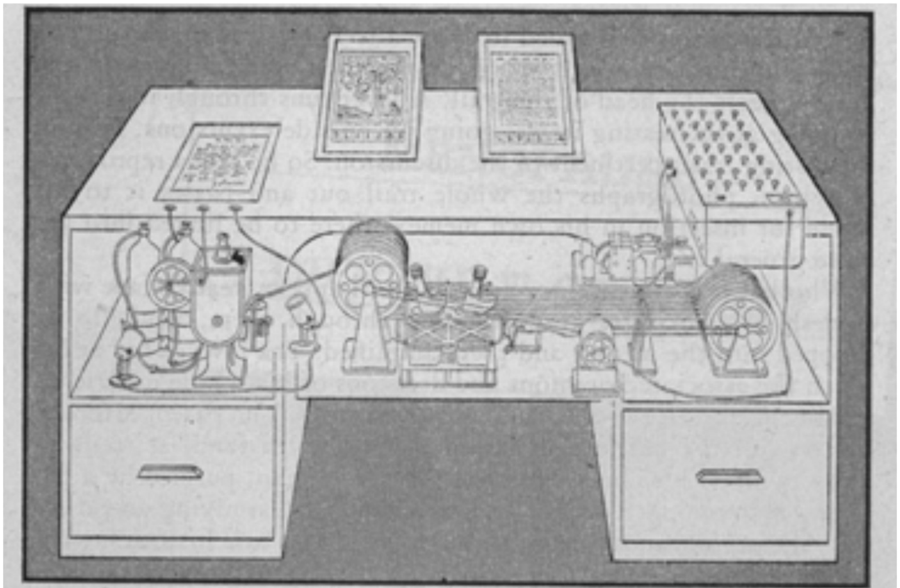
Of these anticipated problems was that of the sharing and storage of information. Bush was aware that the progress being made in science could be hindered by information overload stating, "specialization becomes increasingly necessary for progress, and the effort to bridge between disciplines is correspondingly superficial."[33] The problem he saw was one of scope: "the difficulty seems to be, not so much that we publish unduly in view of the extent and variety of present day interests, but rather that publication has been extended far beyond our present ability to make real use of the record."[34] Science and engineering were producing so much that it was becoming hard for the individual researcher or research team, to not only get through everything but interact with an entire body of literature in real time.[35]

To this end he laid out in fine detail his idea for a machine that would allow for this interaction. He called his proposed device the Memex.[36] The hypothetical Memex was to be a desk-like machine that would have used existing keyboard, microfilm, projection, and magnetic tape technologies.[37] Put together, a person would be able to access anything in their microfilm library such as articles, documents, photos and encyclopaedias. The information would all be

---

[32]Bush, "As We May Think," Introduction.
[33]Bush, "As We May Think," Section 1.
[34]Bush, "As We May Think," Section 1.
[35]Bush, "As We May Think," Section 1.
[36]Bush, "As We May Think," Section 6.
[37]Interestingly the Memex is no longer hypothetical two people from Seattle built one as a project in 2014. Trevor F. Smith, "Memex #001 Demo," YouTube, 5:51, Trevor F. Smith, June 19, 2016, https://www.youtube.com/watch?v=pW4SS_9nXyo; "Memex #001," Trevor F. Smith, accessed July 5, 2019, https://trevor.smith.name/memex/l.

catalogued and pages could be linked. The user could also annotate any of the pages, and share the annotated pages.[38]



MEMEX in the form of a desk would instantly bring files and material on any subject to the operator's fingertips. Slanting translucent viewing screens magnify supermicrofilm filed by code numbers. At left is a mechanism which automatically photographs longhand notes, pictures and letters, then files them in the desk for future reference.

Figure 1: Bush's Memex[39]

This machine was hypothetical but the process and ideas around the ability to store, search, read, link, change, and share information were very influential. In his own words, Bush's machine would have changed the "ways in which man produces, stores, and consults the record of the race."[40] However, computers at this time were still cumbersome and unreliable - they were not yet interactive in the way we know them today.[41] Computers like the UNIVAC worked by producing outputs, a value. This value would be worked with outside the computing environment, on paper or put into a table of values, and then put back in the machine as a

---

[38]Bush, "As We May Think," Section 2.
[39]This illustration was taken from the *Life Magazine* reprint of Bush's article. Vannevar Bush, "As We May Think," *Life Magazine,* vol. 19, no. 11, September 10, 1945, 123.
[40]Bush, "As We May Think," Section 8.
[41] J.W., Cortada, "The ENIAC's Influence on Business Computing, 1940s-1950s," *Annals of the History of Computing, IEEE* 28, no. 2 (2006): 28.

variable to fine a subsequent value. The interaction was in the mechanical and digital programming of the machine. This reality made the Bush article a striking departure from the ideas of computing of the time. People began thinking of themselves as programming pre-written applications, and building computers, rather than working directly with the computers.

J. C. R. Licklider, Man and Machine Working Together:

Bush's vision was not lost on the computing world however. Many would see people and computers acting in tandem to better achieve ends. In 1960 a psychologist and computer scientist by the name of J.C.R. Licklider wrote a paper describing what he called a "man-computer symbiosis."[42] He saw that computers of his day were designed, "primarily to solve preformulated problems, or to process data according to predetermined procedures."[43] Computers of the time were clunky and not intuitively connected to the way people thought or acted. A problem would arise and the computer would have to be programmed for the problem. Once this was done inputs would be made. Then the users would have to wait while the computer worked. This took a lot of patience and would try the willpower of the more proactive of users. Licklider and others could see the possibilities of the computer for human creativity if computers could be made to be more interactive. What he envisioned was "a very close coupling between the human and electronic members of the partnership"[44] What this meant was computers would do the work that people could not, in any field. The stop and go working of computers meant that there was no way to deviate if one wanted the process to change, or had to try alternative methods, they had to stop all processes and reprogram the machine. This put a stop to the creativity of the user, who was forced now to work on a maintenance task rather than a creative one. Licklider's solution

[42]J.C.R. Licklider, "Man-Computer Symbiosis," *IRE Transactions on Human Factors in Electronics,* vol-1 (March 1960): 4-11.
[43]Licklider, "Man-Computer Symbiosis," 5.
[44]Licklider, "Man-Computer Symbiosis," 5.

was to say that a computer needed to work in real time with the user, no waiting, and no maintenance. This would better "bring computing machines effectively into processes of thinking."[45] His hope was that, "in time contributions of human operators and equipment, will blend together so completely in many operations, that, it will be difficult to separate them in the analysis."[46] Licklider's ideas of man-computer symbiosis have largely come true, though he was not the only one to have perceived the value of this outcome.

Douglas Engelbart, At Our Fingertips:

One such person was Douglas Engelbart. In 1945 Engelbart found himself stuck, care of the US Navy, on Leyte Island in the Philippines. There Engelbart came across Bush's article "As We May Think" in a copy of *The Atlantic*. He was struck by the idea of the Memex saying later that "the whole concept of helping people work and think that way just excited me."[47] This excitement would later drive him to an attempt to create "a computer that could interact with the different capabilities that we've already got," referring, for example, to the abilities to do complex calculations quickly.[48] This sentiment came from a need he had felt, during a time of re-evaluation of his life after the war, to make the world a better place. Science was dealing with increasingly more complex problems that required the expertise of more numerous and varied people to solve. The amount of work involved in calculating artillery trajectories, and atomic tests, was becoming burdensome already by the 1940's. Mathematician Harry Polachek writes that "the [US] Army was constantly searching for methods or devices to accelerate the computation or to expand its computation facilities."[49] The work involved in making calculations

[45]Licklider, "Man-Computer Symbiosis," 5.
[46]Licklider, "Man-Computer Symbiosis," 6.
[47]Walter Isaacson, *The Innovators: How a Group of Hackers, Geniuses, and Geeks Created the Digital Revolution,* (New York: Simon & Schuster, 2014), 273; Paul E. Ceruzzi, *A History of Modern Computing,* 2nd ed. (Cambridge: The MIT Press, 2003), 260.
[48]Isaacson, *The Innovators,* 275.
[49]Harry Polachek, "Before the ENIAC." *IEEE Annals of the History of Computing* 19, no. 2 (1997), 25. This was not just happening in the fields of military research. All sectors from the 1940's onward began to see a need for an

was moving beyond the capabilities of the legions of mathematicians that were needed for scientific work. Concerning humans in the twentieth century Engelbart saw that, "the *complexity* of his problems grows still faster, and the *urgency* with which solutions must be found becomes steadily greater in response to the increased rate of activity and the increasingly global nature of that activity."[50] His mind was on how computers could be the vehicle for dramatically improving humanity's ability to deal with this complexity. "If in some way, you could contribute significantly to the way humans could handle complexity and urgency, that would be universally helpful:" all people would be able to benefit from the use and results of such a system.[51]

This line of thought culminated in a 1962 report entitled "Augmenting Human Intellect."[52] In this paper he argued that human capabilities could be coupled, via interactive tools, with computer processing, memory and indexing, to augment human capabilities. This was not some cybernetic utopian thinking but rather, "By "augmenting human intellect" we mean increasing the capability of a man to approach a complex problem situation, to gain comprehension to suit his particular needs, and to derive solutions to problems."[53] Man would use machines as a tool that was created specifically to act in a man-computer symbiosis. The benefits of a more interactive system would be "more-rapid comprehension, better comprehension, the possibility of gaining a useful degree of comprehension in a situation that

---

increase in computational power and accuracy. Keith Smillie writes on how during the 1950's and 60's the University of Alberta's Mathematics, Agriculture, Physics and Engineering departments all benefited by the introduction of the LGP-30. Keith Smillie, "Early Computing at the University of Alberta and the Introduction of LGP-30," *IEEE Annals of the History of Computing,* vol. 29 (2007), 65-73.

[50]Douglas Engelbart, "Augmenting Human Intellect: A Conceptual Framework," SRI Summary Report AFOSR-3223, Prepared for: Director of Information Sciences, Air Force Office of Scientific Research. SRI International, (October 1962), The Douglas Engelbart Institute, accessed July 5, 2019, http://dougengelbart.org/content/view/138. 1a2.

[51]Valerie Landau, "How Douglas Engelbart Invented the Future," *Smithsonian Magazine* (January 2018), accessed July 5, 2019, https://www.smithsonianmag.com/innovation/douglas-engelbart-invented-future-180967498/.

[52]Engelbart, "Augmenting Human Intellect."

[53]Engelbart, "Augmenting Human Intellect," 1a.

previously was too complex, speedier solutions, better solutions, and the possibility of finding solutions to problems that before seemed insoluble."[54] The need and the intention were there to create an interactive system.

Over the next six years Engelbart worked with Licklider, and computer scientist Robert Taylor and others at the Advanced Research Projects Agency (ARPA) labs, to find a way for humans to interact smoothly with machines. Working at the Stanford Research Institute (SRI) with an experimental time-shared Q-32 computer in Santa Monica, the team got to work on various new ways of connecting people to machines.

Figure 2: "The on-line system workstation showing the CRT display, keyboard, pushbuttons, and mouse."[55]

[54]Engelbart, "Augmenting Human Intellect," 1a.
[55]William K. English, Douglas Engelbart, and Melvyn L. Berman, "Display-Selection Techniques for Text Manipulation," *IEEE Transactions on Human Factors in Electronics*, vol. HFE-8, no. 1 (March 1967), Fig. 1. 1c, accessed July 5, 2019, https://web.stanford.edu/dept/SUL/library/extra4/sloan/MouseSite/Archive/AugmentingHumanIntellect62/Display1967.html.

One of the first things that was done was to use a Cathode Ray Tube (CRT) to display data. The screen had been a staple of Engelbart's idea since the beginning. He would state in an interview later that

> I just knew that if a computer could punch cards or send information to a printer, then electronics could put anything you want on the screen. If a radar set could respond to operators pushing buttons or cranking cranks, certainly the computer could! There was no question in my mind that the engineering for that would be feasible, so you could interact with a computer and see things on a screen.[56]

Using a screen to display and input data, rather than paper, shortened the amount of the time needed to get the data. The values would be on the screen: no need to wait for a printer or a card puncher to display them, or to wait while the paper was read by the machine.

This simplified the interaction with the machine and allowed the user to focus on the tasks at hand without having to do multiple secondary tasks. The goal was to create something that anyone could readily pick up and work with. This had the added benefit of giving the computer a face. The CRT acted as a recognisable thing that would act psychologically as the representation of the whole machine, and act as a liaison between the transistors and the user. This would have the effect of making the computer less intimidating through simplification. Nevertheless the hypothetical user still needed to manipulate the device. The SRI team was not looking to just use any device, though. They wanted to "determine the best means by which a user can designate textual entities to be used as "operands" in the different text-manipulation operations."[57] Thus they went about designing and testing various means of manipulating selections on a screen in the simple interactive environment displayed on the CRT monitor.

---

[56]Valerie Landau, and Eileen Clegg, *The Engelbart Hypothesis: Dialogs with Douglas Engelbart* (Next Press, 2009), 31.

[57]English et al., "Display-Selection Techniques for Text Manipulation," 1a1.

The team opted to use a "typewriter-like" keyboard as the method of inputting values such as numbers and letters.[58] The user would still need a way of selecting sets of values or abstractions present on the screen. The best method of selecting objects on the screen was by no means obvious, as they were faced with the question of how to non-virtually interact with a virtual environment. They had a few ideas, most involving a movable proxy on the screen. Engelbart's team designed what become known as a cursor, which they referred to as a "bug," and which the user could manipulate via a control device. Many different types of control devices were built and tested: a joystick, a grafacon, a knee control system, and a mouse. Others represented a different approach such as the light pen. All the options functioned as a selector, but the team was looking for simplicity and efficiency, as well as functionality. As these all represent different modes of interactivity it is instructive to examine their characteristics.

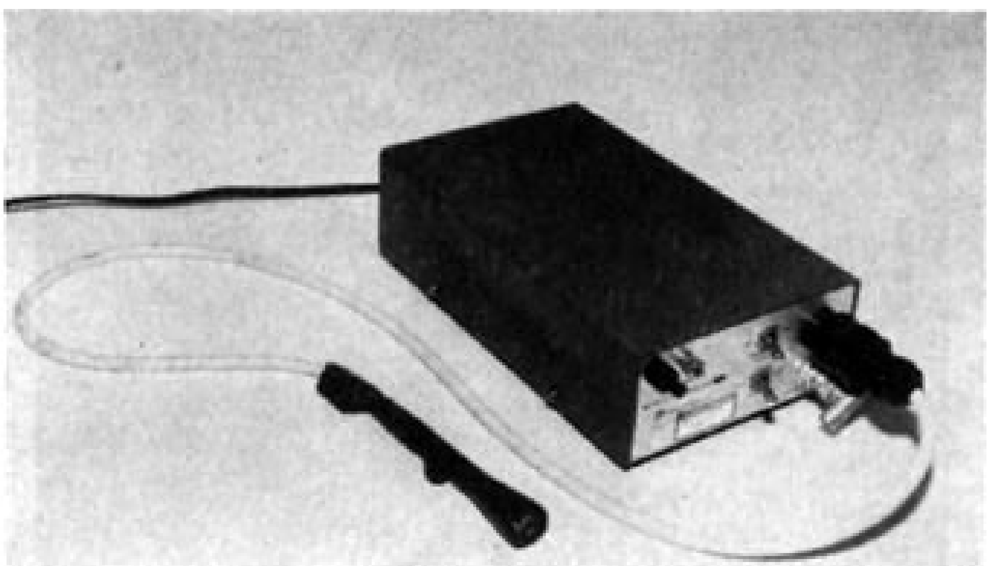


Figure 3: The Light Pen[59]

One device the team tried was a light pen, shown in Figure 3 (above). The light pen would scan the screen from the outside and transmit selections to the computer. It was made of a

[58]English et al., "Display-Selection Techniques for Text Manipulation," 1c1.
[59]English et al., "Display-Selection Techniques for Text Manipulation," Fig. 5. 2f.

stylus attached by a fiber optic cable to a photo multiplier tube. The user used the pen to make selections by pointing the pen "at the desired character on the CRT screen with the aid of a projected circle of orange light indicating the approximate field of view of the lens system."[60] When the orange light was over the desired spot the user could press a button in the pen to make the selection.

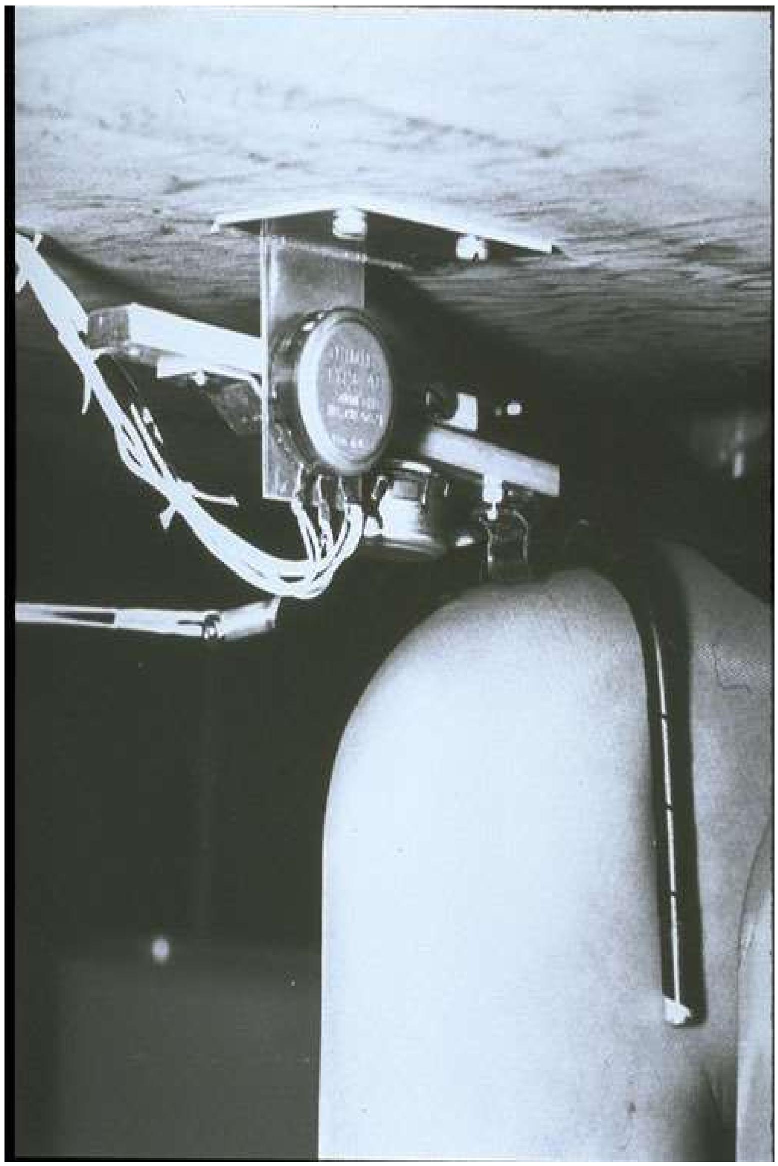


Figure 4: The Knee Control bug-positioning device[61]

[60]English et al., "Display-Selection Techniques for Text Manipulation," 2f3.
[61]English et al., "Display-Selection Techniques for Text Manipulation," Fig. 4, 2e.

One of the more interesting ideas was the knee control system, shown in Figure 4 (above). This mechanism was placed underneath the table above the user's knee. This made it an attractive option as it would have allowed for both hands to remain free for data input on the keyboard and other input devices. The cursor could be moved around by the knee's lateral movements, and selected by pushing a lever with an upward movement of the leg.[62] This system, while interesting, was reliant on the user's knee which as a manipulator is neither as accurate nor subtle as the hands, which are better calibrated for humans to manipulate things seen by the eye.

Figure 5: A woman using the Graficon to control the bug.[63]

The grafacon, shown in Figures 5 and 6, drew coordinates on a field acting as an analog for the screen. It consisted "of an extensible arm connected to a linear potentiometer, with the housing for the linear potentiometer pivoted on an angular potentiometer. The voltage outputs from the grafacon represent polar coordinates about the pivot point...which represent rectangular

[62]English et al., "Display-Selection Techniques for Text Manipulation," 2e.
[63]Christina Engelbart, "Inventing the Mouse," *Historic, Collective IQ Review,* April 12, 2013, https://collectiveiq.wordpress.com/2013/04/12/inventing-the-mouse/#prettyPhoto.

coordinates."[64] Essentially it was an arm that would point on a field representing the screen, attached to two bars, whose movement would indicate the position on screen. To move and select with the cursor, the user would trace out coordinates and the cursor would respond accordingly to the grid on the screen. Depressing the pointer knob would act to select characters on the screen at the coordinates that correspond to the relative position of the pointer knob.

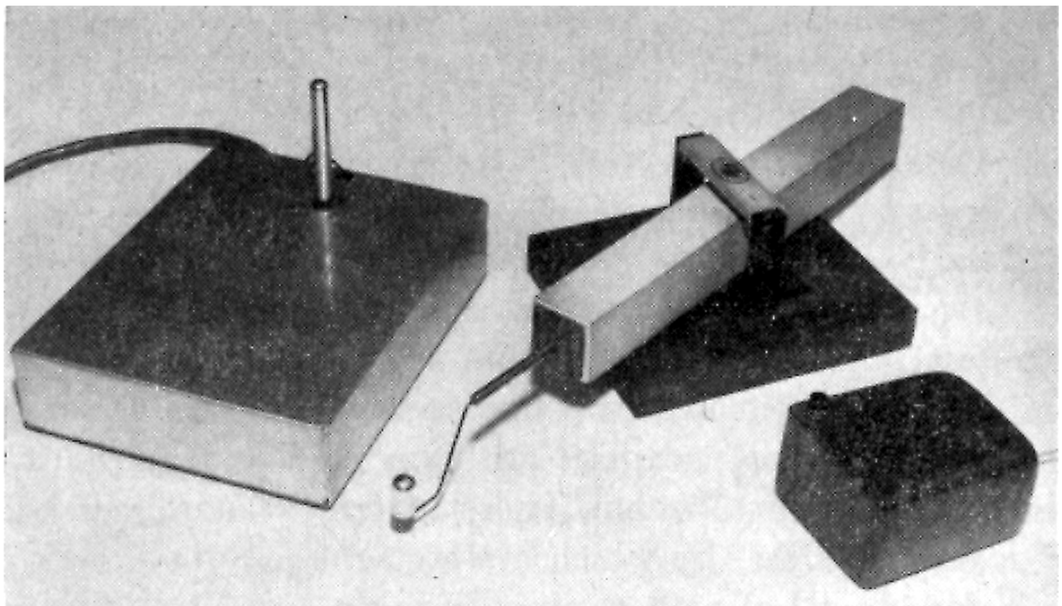


Figure 6: The Joystick (left), Grafacon (middle), and Mouse (right)[65]

The joystick, shown above in Figure 6, was a much simpler design than those that can be found today, being only a small knob affixed to a box. The joystick worked much like the grafacon by being affixed to "two potentiometers, which measured the movement along two axis which were resolved into a signal that could be interpreted on the screen as position data."[66] This device ended up being the least effective when tested by both experienced and inexperienced users.[67] The team attributed the joystick's inefficiency to the way its software was implemented

[64]English et al., "Display-Selection Techniques for Text Manipulation," 2b2-2b2a.
[65]English et al., "Display-Selection Techniques for Text Manipulation," Fig. 2, 2b.
[66]English et al., "Display-Selection Techniques for Text Manipulation," 2c.
[67]English et al., "Display-Selection Techniques for Text Manipulation," 5b4.

as well as the scale. The joystick lacked the range of movement that most of the other control devices had. They made up for this by enhancing the ratio of the joystick's motion and the cursor's movement from 2:1, which the graficon was set to, to 4:1.This required the user to use more subtle movements.

Figure 7: Engelbart's Mouse[68]

The most successful device which team designed was a box that the user could move with their hand. This device, shown above in Figure 7 and below in Figure 8, was nicknamed a "mouse" because of its shape and the tail coming out the back. Engelbart claimed that, "I don't know why we called it a mouse, it started that way and we never did change it."[69] The mouse itself "is constructed from two potentiometers, mounted orthogonally, each of which has a wheel attached to its shaft." Essentially two rollers would be moved by a ball moving on a surface. The rollers corresponded to an axis and the amount of movement indicated the amount of change on the screen position on that axis. Using sketches received from Engelbart, fellow researcher Bill English built the prototype from mahogany.[70] The mouse's design worked well as it "assumed

[68]"First Mouse," Stanford University, accessed July 10, 2019, https://web.stanford.edu/dept/SUL/library/extra4/sloan/MouseSite/Archive/patent/Mouse.html.
[69]"Part 4 of 10: Engelbart and the Dawn of Interactive Computing: SRI's 1968 Demo (Highlights)," YouTube, 1:41, SRI International, December 11, 2008, https://www.youtube.com/watch?v=hRYnloqYKGY&list=PLCGFadV4FqU2yAqCzKaxnKKXgnJBUrKTE&index=4.
[70]Benj Edwards, "The Computer Mouse Turns 40," *Macworld,* (December 9, 2008), accessed July 5, 2019, https://www.macworld.com/article/1137400/input-devices/mouse40.html

that the surface on which the user moves the mouse is not a primary component of the system."[71] Since the user did not have to pay attention to the control surface, the user's focus could remain on the cursor on the screen.

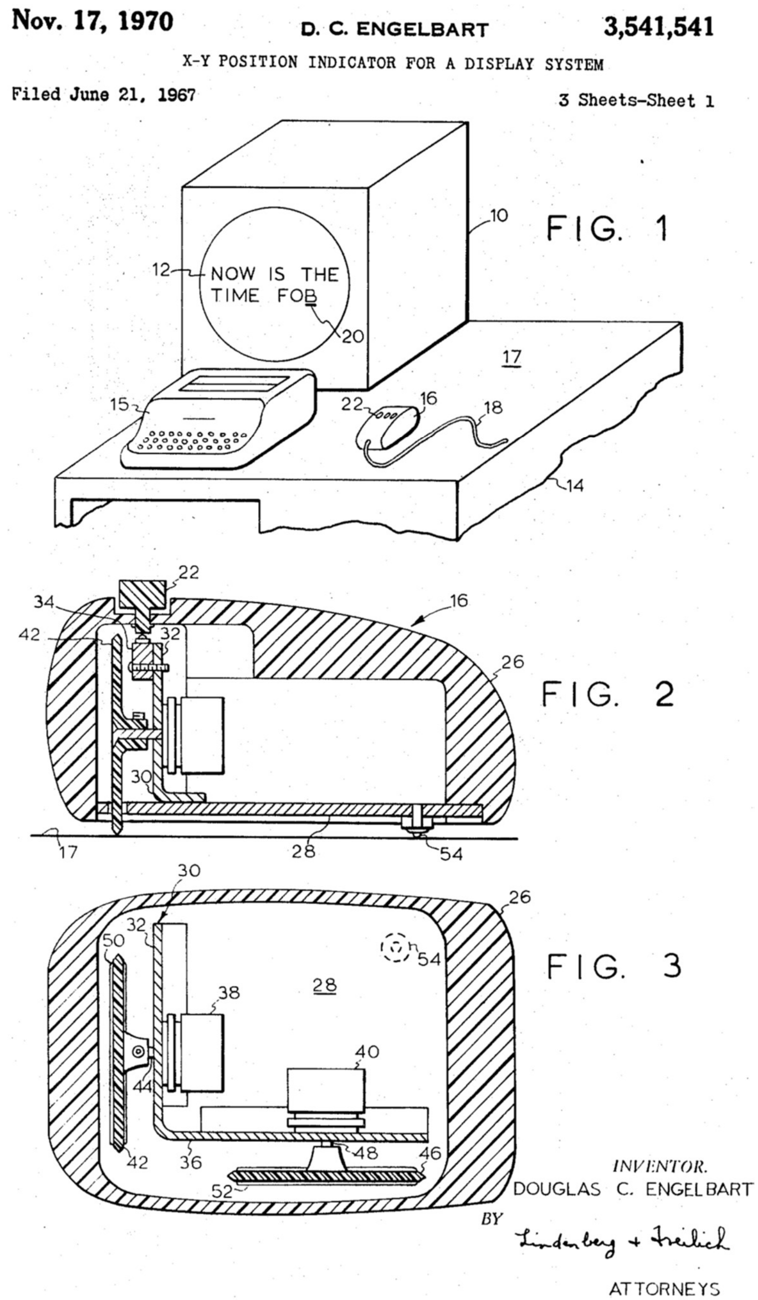


Figure 8: Engelbart's drawing of the mouse from the patent[72]

[71]Thierry Bardini, *Bootstrapping: Douglas Engelbart, Coevolution, and the Origins of Personal Computing* (Stanford: Stanford University Press, 2000), 98.
[72]Douglas C. Engelbart, X-Y Position Indicator for a Display System. US Patent 3541541, filed June 21, 1967, and issued November 17, 1970.

The devices were subjected to tests by allowing experienced and inexperienced users to explore the system and perform certain tasks, most of which involved selecting characters on the screen with the various devices. Subjects familiar with the systems were tested by how fast and accurately they could line up the cursor, the little white + seen in Figure 9 (below), and a specified X on the screen.[73]

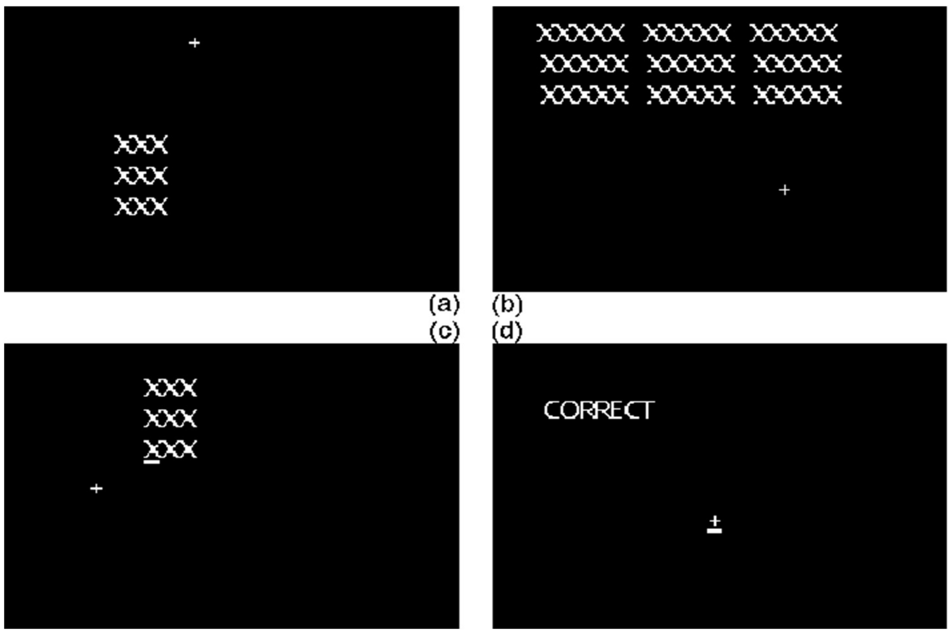


Figure 9: Display screen of test system[74]

They then analysed the data collected by measuring the distance to the targets, over the time it took to get a correct selection of the target character, as well as number of errors in each test.[75] The results showed that experienced users were more adept with the mouse at manipulating the cursor or bug, while inexperienced users tended to perform better with the light pen or the knee control.[76] It was thought that "the light pen exploits one's inherent tendency to select something by straightforwardly "pointing" at it rather than by guiding a bug across a screen toward it from a

[73]English et al., "Display-Selection Techniques for Text Manipulation," 3.
[74]English et al., "Display-Selection Techniques for Text Manipulation," Fig. 6, 3b.
[75]English et al., "Display-Selection Techniques for Text Manipulation," 4.
[76]English et al., "Display-Selection Techniques for Text Manipulation," 5b.

remote control. This means that an inexperienced subject can become reasonably proficient in using a light pen with relatively little practice."[77] The strength of the pen was that it acted as the extension of a finger when pointing and selecting.

As intuitive as was the light pen, the mouse was deemed the "best" by the experienced users, as it caused little fatigue and was satisfying to use. Many of the variables that were deemed to affect the user experience were: where it was reached; how it fit in the hand; the ratio of how much a movement affected the movement of the bug on the screen; how the select button or switch was toggled and how much force was needed to move the device; where the arm and wrist rested; the coordination of the arm to the fingers, and weather the bug stayed put when one removed their control.[78] Having measured and tested the different approaches the team claimed t that "time that it seems unrealistic to expect a flat statement that one device is better than another."[79]

The mouse however became the most recognisable and widespread selection device for most interactive computer needs. The ease of use and accuracy of the mouse has made it a standard tool for over 50 years. It has inherent benefits. The cursor and screen remain un-obscured and the focus of the attention is placed on the cursor moving not as a bug but an extension of one's own will floating around the screen. The mouse, like all other control systems, was not chosen arbitrarily. Any control interface will be dependent on the basic human physiology, the needs of the task and the characteristics of the system being interacted with. For example, players of shooter games will default to the mouse/keyboard interface, while players of fighting games may opt for a controller, and someone using a flight simulator will gravitate to a

[77]English et al., "Display-Selection Techniques for Text Manipulation," 5b3.
[78]English et al., "Display-Selection Techniques for Text Manipulation," 6b3.
[79]English et al., "Display-Selection Techniques for Text Manipulation," 6c2.

joystick. The mouse would prove the more attractive option for the oN-Line System (NLS) that the SRI team was developing. Because of this, the team would go on to use the mouse, toggle (a series of buttons for the left hand), and keyboard interface for their demonstration of their interface and their NLS system in what would become known as the Mother of All Demos.

On December 9th 1968 Engelbart and his team were able to show off their new system and hardware. At the Fall Joint Computer Conference in San Francisco, Engelbart sat at a desk with his keyboard screen and mouse, in front of a standing crowd packed in to see the demonstration. His display and his face were projected on a large screen showing his manipulation of the system located off site at Stanford in real time. He began by asking, "If in your office, you, as an intellectual worker, were supplied with a computer display backed up by a computer that was alive for you all day, and was instantly responsive to every action you have how much value could you derive from that?"[80] This was what the NLS was offering: an interactive system that would compliment any task that you could program it to do. Instead of laying out the specs and workings of the system he showed what the program could do by using the system in front of the audience. He wrote out the word "word" and selected it, copied it, then erased it, and organised the copies into groups that could be collapsed them into discrete units

[80]SRI International, "Part 1 of 10: Engelbart and the Dawn of Interactive Computing: SRI's 1968 Demo (Highlights)," YouTube, 0:38, SRI International, December 11, 2008, https://www.youtube.com/watch?v=VScVgXM7lQQ&list=PLCGFadV4FqU2yAqCzKaxnKKXgnJBUrKTE&index=1. This playlist of highlights, uploaded by SRI, focuses on discrete elements of the NLS systems functions presented at the MOAD. A complete recording of the MOAD can be found here: SRI International, "1968 "Mother of All Demos" with Doug Engelbart & Team (1/3)," YouTube, 34:45, Doug Engelbart Institute, March 12, 2017, https://www.youtube.com/watch?v=M5PgQS3ZBWA; SRI International, "1968 "Mother of All Demos" with Doug Engelbart & Team (2/3)," YouTube, 33:28, Doug Engelbart Institute, March 12, 2017, https://www.youtube.com/watch?v=hXdYbmQAWSM; SRI International, "1968 "Mother of All Demos" with Doug Engelbart & Team (2/3)," YouTube, 32:42, Doug Engelbart Institute, March 12, 2017, https://www.youtube.com/watch?v=FCiBUawCawo. A transcript of the MOAD:"1968 Demo - Douglas C. Engelbart, PhD et.al.," Doug Engelbart Institute, accessed July 7, 2019, http://dougengelbart.org/pubs/video/fjcc68/Englebart's-1968-Transcription.html.

that could be expanded.[81] He brought up his wife's shopping list, by typing in a command that produced a list displayed on the screen. He proceeded to alter the list by selecting various items, of which he changed the order with a couple clicks of the mouse. He added hierarchical levels to the list by selecting a value and adding subordinate values that were hyperlinks to the top hierarchical value. Finally he began scrolling down the list selecting and editing, moving words and adding new ones. This demonstrated a way to interact with the data itself, in such an effortless manner that would be familiar to anyone who has used a computer today, so much so that the gravity of the demo is not immediately apparent.

At this point Engelbart handed the presentation over to Jeff Willison who showed how the SRI team had navigated the NLS system, which had grown to be large for a program of that time.[82] Here Willison did something that would later be taken for granted by most people.[83] He selected a piece of text on the screen with the cursor and was brought to the information that that link referred to. Engelbart elaborates on this explaining what was happening,

> The programming itself represents a very good example for me, where following from the basic philosophy that concepts come in structure and you'd like to structure your data base, or information base that way and have a tool of getting around it. The way we've got our records for programming organized and then the special languages using the hierarchy, using the names of places in there as labels of statements, so both NLS treats them as the name they can jump to, the complier that complies those files we're looking at treats those as the labels for those sub-routines and procedures.[84]

This was a demonstration of the hypertextual abilities that the team had built into the NLS system. The SRI team had build the NLS as a structure of conceptual units that could act in

---

[81]SRI International, "Part 2 of 10: Engelbart and the Dawn of Interactive Computing: SRI's 1968 Demo (Highlights)," YouTube, 6:24, SRI International, December 11, 2008, https://www.youtube.com/watch?v=Xptc6f3Daoo&list=PLCGFadV4FqU2yAqCzKaxnKKXgnJBUrKTE&index=2.
[82]SRI International, "Part 6 of 10: Engelbart and the Dawn of Interactive Computing: SRI's 1968 Demo (Highlights)," YouTube, 3:43, SRI International, December 11, 2008, https://www.youtube.com/watch?v=qisaqiysp2s&list=PLCGFadV4FqU2yAqCzKaxnKKXgnJBUrKTE&index=6.
[83]And rightly so, these systems, though it is good to understand them, are designed so that we don't have to think about the minutiae of our actions, much as we don't think about how it is we take in air or blink.
[84]SRI International, "Part 6 of 10."

reference to each other. Willison stated that they used this code to save time while using time shared computers. This would have been done in order to make every hour count during development. It also aligned with the philosophy behind designing the mouse: decreasing the time it takes to access and interact in order to make the workflow intuitive, unobtrusive, and easy to learn.[85]

Hypertext:

The SRI team could not take credit for the idea of hypertext. Nothing is created in a vacuum, and members of the computer science community were continually interacting with the ideas of other members and groups. This idea of hypertext mirrored Bush's idea of associative indexing, which would have been a hypothetical function of the Memex. The basic idea, Bush writes was "a provision whereby any item may be caused at will to select immediately and automatically another."[86] Bush noted that the mind follows associative trails, a machine would not be as flexible or fast as the mind but, "but it should be possible to beat the mind decisively in regard to the permanence and clarity of the items resurrected from storage."[87] Users could build associative trails or input readymade ones, which could be dropped into the Memex and strengthen the associative web in the Memex's encyclopedia's.[88]

Ted Nelson was working closely with hypertext at this time, and his ideas surrounding hypertext mirrored Bush's ideas surrounding associative trails closely. As a youth while staying at the family farm, Nelson had encountered Bush's article "As We May Think." Like with many others, the ideas surrounding the Memex had planted seeds of thought in Nelson. Later in his life it struck him that Bush's Memex represented "a way of thinking," that "of being able to follow,

[85]SRI International, "Part 6 of 10."
[86]Bush, "As We May Think," Section 7.
[87]Bush, "As We May Think," Section 6.
[88]Bush, "As We May Think," Section 8.

not just the exposition of a teacher, and I didn't like teachers very much, but being able to follow the thought processes of a researcher discovering something, as Bush points out in "As We May Think."" Nelson, frustrated with hot media, like lectures, gravitated to cooler media such as comic books and 3-D movies, and looked to develop ways in which one could arbitrarily travel, "through the new space of the media of the heart and mind." Nelson was referring, in part, to the digital creative content that humanity would create as a result of the human-computer symbiosis. He saw that, "the future of personkind would be at the computer screen," and it was important to have a system that would allow for it to be cool rather than hot, for it to be participatory rather than merely passive content absorption. A cooler medium would be more accessible and better for a mass creativity. One highly skilled person has the ability to make something that everyone can use. That scenario only allows for the creativity of a few. Having many people acting creatively together would produce more results with greater diversity. An intensification of the medium-message cycle, available to all through this new media, "would fuse into a new combination of movie and interactive thingamabob." The interactive machines that were emerging in the mid 1960's had the potential to do this but complexity was still a problem. [89]

Figure 10: A normal narrative found on paper[90]

[89]All the quotes from this paragraph taken from: Ted Nelson, "Ted Nelson – Vannevar Symposium 1995," YouTube, 41:30, Colin McDonnell, July 13, 2015, https://www.youtube.com/watch?v=A3mptA5sZTM.
[90]Theodor H. Nelson, *Computer Lib: You can and must understand computers now / Dream Machines: New freedoms through computer screens—a minority report* (1974), 44.

People were used to, and still are, interacting with digital objects as if they were paper, sequentially reading and scrolling. Nelson argued in the early 1970s that, “The structures of ideas are not sequential. They tie together every whichway. And when we write, we are always trying to tie things together in non-sequential ways.”[91] Nelson had come up with the term “hypertext” in the mid 1960s’ when trying to describe this.[92] “By “hypertext” I mean non-sequential writing.”[93] This he laid out in multiple ways in his book *Computer Lib / Dream Machines* in 1974, where he would elaborate on his radical vision of computer usage. Bush noted that to realize these ideas Nelson committed himself to the Xanadu project, which was an attempt to create the first truly hypertexual digital environment.[94]

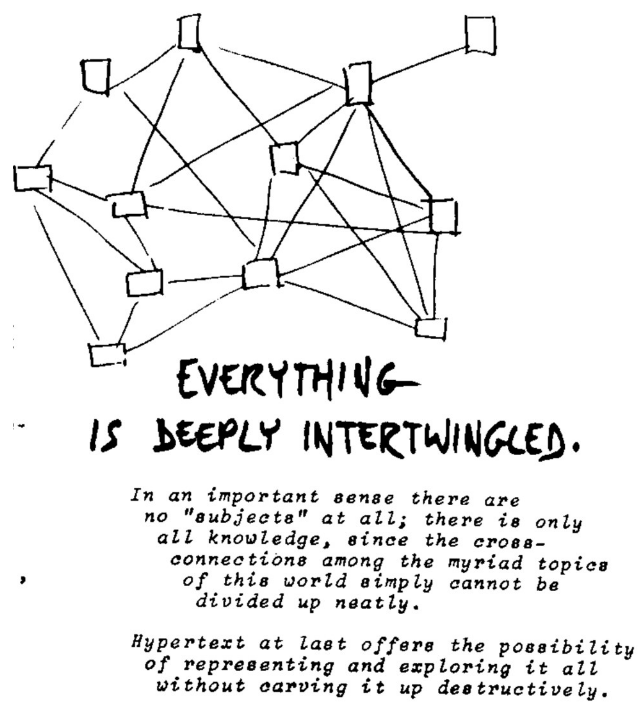


Figure 11: A hypertextual nonsequential narrative[95]

[91]Nelson, *Computer Lib/Dream Machines,* 44.
[92]Campbell-Kelly et al., *Computer,* 234.
[93]Nelson, *Computer Lib/Dream Machines,* 44.
[94]Project Xanadu, “Project Xanadu Founded 1960 The Original Hypertext Project” accessed July 5, 2019, http://xanadu.com/; Ted Nelson, “Ted Nelson Demonstrates XanaduSpace (by Arthur Bullard),” YouTube, 7:31, TheTedNelson, July 10, 2013, https://www.youtube.com/watch?v=1yLNGUeHapA.
[95]Nelson, *Computer Lib/Dream Machines,* 45.

Nelson worked on Xanadu for thirty years, but ultimately his project would fail to gain any meaningful support. Hypertext however would become a staple of interactive computing environments. Nelson envisioned a completely open and interconnected environment that would not rely on the characteristics of outdated mediums. The NLS system, for instance, was designed with a QWERTY keyboard and displayed information in the same manner as a piece of paper would. This was intentional, as each medium plays off of previous medium; people in offices had been used to using typewriters for more than half a century and would find the system somewhat familiar.[96] Thus instead of a purely open environment where everything was interwingled, as Nelson called it, the NLS and successor systems were display mediums that had hypertextual elements. The content is displayed on the screen, but the ability to select and navigate interactively with discrete units of data, a page or a file, became a staple of modern computing environments.

These ideas had been taken to heart by computer scientist Alan Kay. Kay's interests lay in the invention of a personal computer that was portable and could be used by children.[97] He was concerned with the computer as a tool for the communication of ideas, but he saw that there were a few difficulties. The principal problem was the understanding of the medium. "Anyone who wants to receive a message must first internalize the medium so it can be subtracted out to leave the message behind."[98] People were not used to using computers, and he needed the interface to be made so that computer use would be intuitive from the first encounter. He began to think about how a GUI would work as a medium that would transmit the message. Kay

---

[96]Campbell-Kelly et al., *Computer,* 21-22.
[97] Susan B. Barnes, "Alan Kay: Transforming the Computer into a Communication Medium," *IEEE Annals of the History of Computing,* vol. 29 issue 2 (April-June 2007): 20; Isaacson, *The Innovators,* 287.
[98]Barnes, "Alan Kay," 20.

suggested that, "doing with images makes Symbols."[99] Instead of directly interacting with the code, text operations could be represented with graphical icons.[100] The system would be object oriented. Hypertext did not necessarily need to be textual: it could link an object, the symbolic representation of a process or function. It could be anything.[101]

The idea Kay had was for a small tablet machine that would educate children using an intuitive and interesting interface. His thought was that computing was the new medium, and for any medium to have an effect (a message) it needed to pervade the society in which it worked.

> The ability to ''read'' a medium means you can access materials and tools created by others. The ability to ''write'' in a medium means you can generate materials and tools for others. You must have both to be literate. In print writing, the tools you generate are rhetorical; they demonstrate and convince. In computer writing, the tools you generate are processes; they simulate and decide. If the computer is only a vehicle, perhaps you can wait until high school to give ''driver's ed'' on it—but if it's a medium, then it must be extended all the way into the world of the child.[102]

So easy a child could use it was one of the goals of Kay's Dynabook. Such a system would help to make people literate with the tools that, he felt, would most augment human creativity, just as literacy had done with the written word.

Kay's Dynabook would never be built; however others had seen the potential of Kay's proposed system, and as a result Kay began work at Xerox Palo Aalto Research Center (PARC) in 1973.[103] There, Kay began working with a team developing a personal computer: the Xerox Alto. This proto-PC would incorporate many of the design elements seen in Engelbart's NLS. It had a rudimentary screen for display. A keyboard and mouse for interaction with the machine. The display was designed with an elongated screen to mimic the dimensions of an 8 1/2 / 11

---

[99]Barnes, "Alan Kay," 24; Alan Kay, "User Interface: A Personal View," in *Multimedia: From Wagner to Virtual Reality,* ed. Randall Packer, and Ken Jordan, Expanded ed. (New York: W. W. Norton & Company, 2002), 128-9.
[100]Barnes, "Alan Kay," 23.
[101]Barnes, "Alan Kay," 26.
[102]A.C. Kay, "User Interface: A Personal View," 125.
[103]Campbell-Kelly et al., *Computer,* 260.

piece of paper. The Alto's computing environment though similar in concept, had expanded on what the NLS system had been. Like Kay's proposed Dynabook, it was object oriented and featured icons representing folders and files, similar to what most computer users are familiar with today. [104] These objects were presented in a bitmapped display. Every pixel on the raster scan display was "accessible as a bit of memory."[105] This meant that each bit or pixel could be interacted with in a dynamic way, and pixels could be combined to make images or symbols that could be used to link hypertextually to processes. Users of the Alto could run games, store programs and files, run a word processor, and edit graphics.

Figure 12: The Xerox Alto[106]

While the Alto was a step forward in the evolution of computing it was a commercial failure, with little over 2100 being made of both models.[107] Many of the ideas the Kay and his team brought together were later refined and perfected by Apple and Microsoft in the late seventies and early eighties. Steve Jobs and a group of engineers from Apple visited

---

[104]Campbell-Kelly et al., *Computer,* 260.
[105]Thomas A. Wadlow, "The Xeorx Alto Computer," *BYTE Magazine*, vol. 6 no.9 September 1981, 58, Internet Archive, accessed July 5, 2019, https://archive.org/details/byte-magazine-1981-09/page/n59.
[106]Wadlow, "The Xeorx Alto Computer," 58.
[107]Campbell-Kelly et al., *Computer,* 260.

XeroxPARC in 1979, and subsequently incorporated much of what they saw, including the mouse, windows, and icons, into the development of later Apple computers.[108]

What this means to an archivist is that hypertext will be a property of virtually all software. Like a printout of a spreadsheet, the removal of the hypertextual nature from a record would be to remove most of the record, essentially heating up the record by removing most of the participation. The ability to move from one place to another along the representative associative link is an essential part of a user's experience of that site. A webpage is part of that meaningful communication, and (more than the sum of its parts) it is not enough to look at the source code or the rendered page. Webpages are built deliberately based on their creator's needs, whims, and desires, and were designed with interactivity in mind: the movement and evolution of data using hypertextual means. Though the user can take screenshots of a webpage, this removes much of the background code and user experience by reducing it to a display medium. The Web is a very cool place. Melting it would remove the ice that so characterises its landscape.

Browsers: Connected but We Need Somewhere to Meet

From the "Mother Of All Demos" in 1969 to the birth of the first webpages computing would evolve and personal computing would become a norm. During this time computer scientist Tim Berners-Lee was working as a contractor at the European Organization for Nuclear Research (CERN). Needing a way to keep track of the programs running, the people involved, and any equipment, along with their relationships, he built Enquire. This program acted somewhat "like a card index, but with links between cards."[109] Enquire would have allowed all

---

[108]Isaacson, *The Innovators,* 363-64.
[109]Tim Berners-Lee, "A Brief History of the Web," World Wide Web Consortium, 1994, accessed July 5, 2019, https://www.w3.org/DesignIssues/TimBook-old/History.html.

information in a network to be accessible in the same virtual space.[110] Enquire was never formally adopted, but was used as a tool for Berners-Lee "to keep track of the modules, the users, the documents and everything I needed to note down."[111]

Though it would never become more than a prototype Enquire did give Berners-Lee an idea of what would be needed to make such a system viable:

> There was clearly a need for something like *Enquire* but accessable to everyone. I wanted it to scale so that if two people started to use it independently, and later started to work together, they could start linking together their information without making any other changes. This was the concept of the web.[112]

In 1990 Berners-Lee began developing the Hypertext Transfer Protocol (HTTP) using the capabilities of the UNIX based, NeXTstep operating system.[113] The protocol would act as a standard language that would allow computers to talk to each other easily over a network. The standardized data needed to be rendered into something, so a user could see and interact with the hypertextual content. For this he created a browser program called *WorldWideWeb* (WWW); and to create and edit content for the WWW, he wrote a formatting language called the Hypertext Markup Language (HTML).[114] With HTTP, WWW, and HTML, Berners-Lee had created an interactive system. This system reflected his idealistic views, with which he envisioned information being a democratic tool, with as many people as possible benefiting, as well as collaborating, in an open system.

He saw it as a tool to allow humanity to grow, as more people would be able to create freely in a system with as few constraints as possible.[115] One computer augments a person's

[110]Tim Berners-Lee, *Weaving the Web: The Original Design and Ultimate Design of the World Wide Web by its Inventor,* (San Francisco: Harper San Francisco, 1999), 16.
[111]Berners-Lee, "A Brief History of the Web," Conception.
[112]Berners-Lee, "A Brief History of the Web," Conception.
[113]Tim Berners-Lee, *Weaving the Web,* 28.
[114]Tim Berners-Lee, *Weaving the Web*, 28-29.
[115]Tim Berners-Lee, *Weaving the Web,* 1-2.

natural capabilities, but a virtual space made-up of networked computers could augment the abilities of anyone with access. Not only this, but they would be the creators as well. A person would be able to make their own webpages and show whatever content they wanted in whatever arrangement they desired. They would also be free to build on it and create their own tools. They would also be able to link to anything else on the web, bringing people's creativity into contact.

With the web protocols established, many others saw just as much promise as Berners-Lee. People began to develop browsers in which you could explore hypertext environments in a visual medium. Browsers like Mosaic and Netscape made the networked information easy to access, using the keyboard and mouse layout that had come to be standard for most machines. HTML was not difficult. It had become easy, and often fun, for anyone to make a webpage, when before text based browsers like Lynx had a slight learning curve and were not as inviting.[116] People were beginning to create and interact with each other in these graphical object oriented environments in large numbers. As the web evolved languages like CSS, JavaScript, and Flash were created. These tools changed the way people used the web and computers.

Berners-Lee was playing off of something that had become standard in the world of computer usage: a standardized interface system for most computer environments and needs. This was based on ideas and hardware developed by the SRI team and others. The keyboard and mouse controlling a GUI would become almost universal by the mid-1990s. Engelbart's so-called "mother of all demos" was viewed by many enterprising individuals, many of whom had

---

[116]Lynx is still receiving periodic updates. The last of which was released July 8, 2018, and can be found here: "Lynx2.8.9," Thomas Dickey, last updated July 8, 2018, https://lynx.invisible-island.net/lynx2.8.9/index.html.

also read "As We May Think," and sought to improve or elaborate upon what they had seen and read.

Conclusions:

Much can be said about the development of computers from there. However, this narrative was meant, not only as a way for those not familiar with the history of the interactive tools, that we use to see the way certain properties of their digital records have been constructed, but also to show that they are nested in the histories of their creation and use. This narrative acts as an example and an encouragement to explore other narratives of interactivity. The reader may have noted exceptions, and omissions of examples of technology that are interactive or relevant to the history of interactivity. Mobile devices, flat screens, touch screens, voice recognition, Wi-Fi, and virtual reality, are all examples of technologies that make up part of one's interaction with technology. I have only given the broadest scope of what is meant by, and what is important about, using these tools. All of these devices have had some effect on our society.

These media all represent the way we as a society have come to interact with each other. This is their message. Hot or cool they were not just created arbitrarily. The way we envision our needs and desires intersects with our philosophies, which all inform what systems we want. But we work in a universe which limits us to the possible. The limits and nature of human physiology and psychology predetermines certain outcomes. This means broadly that our visual sense, and tactile manipulation, along with the conceptually associative way we think, leads us to design systems that will work with the way we are constructed. Within these human parameters we can find that there is still infinite diversity, just as user interfaces with the same underlying construction can manifest in countless ways.

Because of the way we have constructed these systems, it is easy to just assume that one will be able to interact with a computer in a certain way, that actions such as scrolling or clicking will have a particular effect. This is not a moral judgment as these systems were intended to feel as natural to use as possible.

The archivist however must go beyond this, and make an attempt to understand the parameters of the system, in order to know the value of the variation of form and function that a digital record represents. We know about books and scrolls. Do we all know why we use a mouse, and what happens when the mouse is subtracted from the interaction with a spreadsheet, a program, or a webpage? Will we know what a mouse is in 100 years, and still be using them? Without documenting our current computing environment something is lost. Understanding the narratives surrounding the GUI and a hypertextual record will allow the archivist to make a better decision as to why a record's interactivity should be maintained.

This is no small task as computers have become ubiquitous in people's everyday lives. Mobile devices connect millions of people from their hands, while giving them the abilities of a dozen more specialised devices. Desktop computers and laptops are a normal part of any residence or workplace, but they lay in more subtle places. People interact with them when they open their fridges, adjust the temperature, and start their car. Each machine is someone's (or more likely some group's) creation, which in turn enabled and informed the actions of those interacting with them. Archivists will need to adapt to the already immense and still growing interactive landscape. As with all records, not all interactive records will need to be kept. Those that are, though, should receive the treatment that they deserve. For this the archivist must work with a thermostat, to make sure that their archive stays cool.

Chapter Two:
Another Tool on the Belt: Fitting Interactivity into Current Archival Contexts

In the previous chapter, I argued that interactivity is an aspect of every digital record and should be maintained if that ascpect of the record is as Marshall McLuhan would call it "cool," that is where more effort is needed to engage with the medium. This notion of "interactivity" was defined with reference to an archival setting as a property of a digital record, where a user's input is integral for its use and comprehension, and where a resulting output, culminating in a set of data to be consumed, is not the primary significance of the media. Essentially all records are interactive, and it is necessary to maintain that interactivity in records where there is a greater degree of user input or participation. It is not hard for many to understand the value of having the interactive elements of a record present for those wishing to engage with the record. Archival scholars such as Jeff Rothenberg have been arguing for similar such initiatives for more than two decades.[117] The main method of preserving interactive digital records is emulation. In this chapter I will go into more detail as to what emulation is and how it compares to other methods of preservation such as format migration, in order to show why emulation is the most promising technique available for preserving interactivity.

More important however, is how it would fit within archival practices and approaches that are available. Emulation is not an easy task, and it will have to either blend into existing archival workflows or become part of prospective archival and information management approaches. With regard to archives, there are two dominant approaches which fall within the scope of this work: the life cycle model and the continuum model. The life cycle model, which T.R. Schellenberg's work made a reality, has been implemented in various forms since the mid

[117] Jeff Rothenberg, "Ensuring the Longevity of Digital Information," Rev ed. (1999): 1-18.

twentieth century.[118] Interactivity could be maintained in such a model, though there would be need for more negotiation between parties, which would be forced to take steps that have become redundant by current technologies, such as fast data transfer, metadata logging, database systems, and further technologies that will be discussed in Chapter Three. The Continuum model, which comes out of Australian schools of thought, emphasises the history of the record and the usefulness of a record across its entire existence, and represents a more streamlined approach in which emulation can be integrated from creation of the record.

Part One: Emulation

Invariably the preservation of digital interactivity becomes an issue of digital preservation and providing of access with that interactivity intact; both problems are mutually inclusive. An archivist looking at a digital record must ask both questions: How do we preserve interactivity, and how do we make the records accessible in an interactive fashion while maintaining the integrity of the records?

The most obvious answer would be to use the original hardware and software. However, this is a short-term solution at best. There are many institutions that preserve old technology and software. The University of Michigan Library's Computer & Video Game Archive, and the University of Illinois Undergraduate Library's Gaming Center are two active ones that collect hardware and software for use. This use is not limited to research; they are available for recreation, and are geared towards getting students interested in video game creation and scholarship.[119] While a useful resource to have at any institution attempting to engage students in

[118] T.R. Schellenberg, *Modern Archives: Principles and Techniques* (Chicago: The Society of American Archivists, 1998).
[119] "About the Collection," University of Illinois Undergraduate Library, accessed July 5, 2019, https://www.library.illinois.edu/ugl/gaming/about; "Computer & Video Game Archive," University of Michigan Library, accessed July 5, 2019, https://www.lib.umich.edu/computer-video-game-archive; The Personal Computer museum, in Brantford Ontario, hosts an impressive collection, though it is closed at the time of writing. "Home," Personal Computer Museum, accessed July 5, 2019, https://pcmuseum.ca/index.asp.

computer engineering and scholarship, these resources are informed by their construction. The University of Illinois collection acts as a library, while the University of Michigan acts as a museum, though they use the term archive. The software and hardware that both these institutions collect and maintain are examples of programs and hardware that were in use. These examples exist without any logical connection to the rest of the collection other than the properties they share as examples of that medium. This is as opposed to a video game that was present in a collection as the result of the actions of a creator in the course of their life, and exists as a logical part of a whole set of records. This is merely the difference between an archive, a library and a museum, but the intention of the institution is important. One of an archive's jobs is to maintain the integrity of the record. Non-archival institutions such as museums and libraries often maintain records with no regard to provenance or original order, as there is no need to for their purposes. The records are kept as examples of a work, rather than as the documents and other things people make as a result of the their daily actions.

While access is achieved by simply using the original systems, the larger issue with these methods is that of preservation, which in turn complicates access. Computer hardware becomes inaccessible through obsolescence faster than paper records become inaccessible through decay. Maintaining a piece of hardware indefinitely is problematic for two reasons. The first problem is cost and expertise. As the majority of the population only has familiarity with computer systems as users, and since the functionalities and affordances of computer systems are changing constantly, the older the computer system the less likely it will be to find someone with working knowledge of that system. Thus the reliance on technology specialists becomes more necessary the older the machine becomes. A computer's components are physically degrade over time, as can be seen in the yellowing of old plastic cases. Hardware also weakens through use, wires and jacks loosen from repeated plugging and unplugging, and solder comes apart from too much cooling or heating.[120] These machines cannot be made to last forever, to try would stress the budgets of any archive would likely increase the neglect, hastening the inevitable.

The second issue is again concerning record integrity. If a piece of equipment is appraised as worthy of keeping in the archive, its use over time would act to damage this particular kind of record (i.e. physical hardware). The practice of making a preservation copy, as with software, so as not to harm the original, cannot be implemented as copies of hardware would be impractical to make, given that parts and expertise are often no longer available.

I have left out a number of practices and examples, but I believe I have made the point clear that keeping the hardware of legacy systems is not a viable strategy for preservation or access. There remains a need to maintain accessible copies of digital records, while maintaining the interactive elements. Unfortunately the answer to this problem for digital records in many archives at present is the same as for paper media. As Michael Fostrom notes, “the digital content has not been appraised prior to acquisition, and the [digital] media is part of a collection consisting chiefly of paper-based materials.”[121] Disks and drives end up filed in acid free folders, with the hope that one day there will be the money, expertise or will to perform digital preservation – and with a prayer that these digital records can even be found in the vaults, and that they are not corrupt when they are finally put through preservation. Often born digital media such as emails, or word processing documents, are just migrated to paper and accessioned as a paper record. However as Terry Cook had already noted in 1995 concerning digital records:

---

[120]Patricia Galloway provides a good anecdote about attempting to revive a Kaypro II in order to access digital materials. She suggests the use of documentation strategies to allay the lack of working knowledge. 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): 60-74.

[121]Michael Forstrom, “Managing Electronic Records in Manuscript Collections: A Case Study from the Beinecke Rare Book and Manuscript Library,” *American Archivist*, vol. 72 (Fall/Winter 2009), 461.

> The content, structure and context of the record change significantly from that of the paper world. The only approximate match with paper is the content element, where the letters and numbers look much the same on the computer screen as on paper. But the structure and especially the context of electronic records are not apparent when retrieved from the text only.[122]

Managing digital records through "print to file" policies mitigates some challenges associated with digital records by ignoring or changing the structure and context of the record. The perceived equivalence of a paper document, and "a "document" on the monitor," has a basis in the construction of these mediums.[123] As discussed in the last chapter computer interfaces were designed with the previous medium in mind, even utilizing the QWERTY layout of the typewriter. This is consistent with McLuhan's observation that the content of every new medium is the medium that preceded it.[124] Cook noted that archivists were receiving "forceful executive orders that all computer-generated information shall be so printed and filed." The understanding of the paper medium made the physical copy carry more authority and presence, which made this migration to another media seem logical. However intentioned, Cook argues that at best, "print-to-paper" is a very short-term band-aid solution that archivists and information managers should be working hard to leave behind them."[125] Greg Bak argues that methods like these, by removing digital records from their native digital environments, dissociate them from the rich metadata that allow digital archivists to demonstrate their provenance and authenticity.[126] He further notes the problematic character of these practices when he asks, "what would stop someone from simply changing (or re-creating) the digital original, reprinting it, and substituting the falsified

---

[122]Terry Cook, "It's 10 o'clock: Do you know where your data are?" *Technology Review* Vol. 98 Issue 1 (1995), 51.
[123]Terry Cook, "Electronic Records, Paper Minds: The Revolution in Information Management and Archives in the Post-Custodial and Post-Modernist Era," *Archives & Social Studies: A Journal of Interdisciplinary Research* vol. 1, no. 0 (March 2007), 422.
[124]Marshall McLuhan, *Understanding Media: The Extensions of Man* (Cambridge: The MIT Press, 1994), 9.
[125]Cook, "Electronic Records, Paper Minds," 422-423.
[126]Greg Bak, "Not Meta Just Data: Redefining Content and Metadata in Archival Theory and Practice," *Journal of Archival Organization*, 13:1-2, (2016), 2-18.

record?"[127] As stated in chapter one, changing the medium too much in favour of just the content changes the record. Filing digital records with the paper documents is not digital preservation and printing content to paper destroys any interactivity.

The only ways to properly maintain the integrity of digital records in an archival environment is through the practices of migration and emulation. The more promising of these two, with regard to interactivity, is emulation. This is the practice of using a computer system to simulate the environment of another hardware or software environment in order to run files that would otherwise be incompatible with the user's computer system. This has the effect of maintaining the look and feel of the record, as well as the bit level data, and thus represents the best way we have now to maintain digital interactivity. While popular in gaming communities as a way to play old games, emulation has also been a useful tool in computer science. Despite this emulation has not been brought into wide use by the archival community.

The cost and complexity of emulation has driven most archives to embrace format migration as the preferred tool for digital preservation. The SAA defines format migration as, "The process of converting a data from an obsolete structure to a new structure to counter software obsolescence." Format migration "may involve changes in the internal structure of a data file to keep pace with changing application versions, such as migration from Word 95 to Word 2000."[128] Often format migrations can be more extreme than this, as with spredsheets and word processor files converted to PDF, or the migration of videogames to video, by recording gameplay with a video camera, as discussed by Guttenbrunner et al.[129] Adrian Brown's overview

[127]Greg Bak, "How Soon Is Now? Writings on Digital Archiving in Canada from the 1980s to 2011," *The American Archivist* vol. 79, No. 2 (Fall/Winter 2016), 296.
[128]"Format Migration," Society of American Archivists, accessed July 5, 2019, https://www2.archivists.org/glossary/terms/f/format-migration.
[129]Mark Guttenbrunner, Christoph Becker, and Andreas Rauber, "Keeping the Game Alive: Evaluating Strategies for the Preservation of Console Video Games," *The International Journal of Digital Curation*, Issue 1, vol. 5 (2010), 80-82.

of the current orthodoxy in digital preservation sees format migration as a viable strategy to keep the content of records accessible ahead of the march of hardware and software obsolescence. He defines format migration as a way to, "transform the original object to a form that is no longer reliant on obsolete technology."[130] John Garret describes migration more succinctly as "a set of organized tasks designed to achieve the periodic transfer of digital materials from one hardware/software configuration to another, or from one generation of computer technology to a subsequent generation."[131] These definitions focus on the essential practice of taking a file in an outdated format, and altering the underlying data so that the content will be useable or accessible on a different or more current medium. The archivist takes the bit level data and translates it using available tools into a format that is readable by the new software and hardware. For example, many of us have written a document in Excel 5.0 only to open it years later to find that it does not translate well, and in fact generates errors and omissions, in Excel 2016. These errors can result in glitches that may render the record, at worst, unreadable or change the look and feel, and possibly even the content of the record.

There is more than one type of migration. One such method is media migration, which is an essential tool of digital preservation, which the SAA defines as "The process of converting data from one type of storage material to another to ensure continued access to the information as the material becomes obsolete or degrades over time."[132] Media migration merely involves copying stored data from one storage media to another, for example from a CD to a hard drive. This does not change the bitstream and the data remains the same. Media migration represents a

---

[130]Adrian Brown, *Practical Digital Preservation* (Facet, 2014), 209.

[131] Task Force on Archiving of Digital Information, Commission on Preservation Access, and Research Libraries Group. *Preserving Digital Information : Report of the Task Force on Archiving of Digital Information*. Washington, D.C.: Commission on Preservation and Access, 1996, iii.

[132]"Media Migration," Society of American Archivists, accessed July 5, 2019, https://www2.archivists.org/glossary/terms/m/media-migration

essential digital preservation strategy, despite its limitations.[133] Though the data is safer on the new media it is not likely to work on progressively newer systems. This is why many migration projects opt for format migration in addition to media migration, so as to have it work with newer systems. This may not be possible as the new formats may be too dissimilar, or the type of digital object may no longer have any analogue that works with current systems, such as a save file from an old game console. The essential problem with format migration is that the underlying data is being changed. While the content of a video or text file may remain consistent from the original format to the next any underling bit data is utterly changed and all metadata may be lost or changed. The record may no longer be reliant on the obsolete technology, but once migrated it becomes reliant on the present technology which is in itself ephemeral.

This change in the underlying data that resulted from format migration, was noted by Geoffrey Yeo who raised the question, “If we have to migrate records to ensure their long-term accessibility, how far, if at all, can we trust that the migrated record is the same as its predecessor?”[134] In asking this he makes two points. The first being that the migrated record is in itself a different record from the original, with its own characteristics and information that it contains. The second point he is making is that without the original there is no way to ensure that the migrated version is representive. Without a working original there is no way to ensure that the functionality and data, the look and feel, would be representive of that original. Emulation would allow the original data to run on newer systems. Though he saw the value of emulation,

[133]David Rosenthal sees digital preservation as an ongoing process requiring prepetual media migration and data auditing. David SH. Rosenthal, "Bit preservation: A solved problem?," *International Journal of Digital Curation* 5, no. 1 (2010), 140.
[134]Geoffrey Yeo, "‘Nothing is the same as something else’: significant properties and notions of identity and originality," *Archival Science* 10, no. 2 (2010), 86.

Yeo noted that emulation is "rarely straightforward," and "at present, for many archival institutions, migration may be the only practically affordable solution."[135]

Though Yeo's pessimism is well founded, there are those who had been looking at emulation as an archival tool for more than a decade when Yeo asked that question. One of the earliest advocates for emulation was Rothenberg, who holds the position that emulation represents the best solution to the question of long-term digital preservation. The quick pace of change in computing technology renders hardware and software obsolete and often unreadable by present computers.[136] He argues that media and format migration represent "an ongoing effort: future access depends on an unbroken chain of migrations with a cycle time short enough to prevent media from becoming physically unreadable or obsolete before they are copied. A single break in this chain can render digital information inaccessible—short of heroic effort."[137] Archives will continue to have to migrate the data from this legacy media *ad infinitum* as computer technology evolves. Rothenberg's position that emulation is a better alternative is based around letting computers evolve around the emulated record. As archives will be upgrading their computers as a matter of course anyways, why continually be upgrading the records' data as well?[138] Rothenberg's point makes sense as when an archivist migrates a record they are continually moving further and further away from being able to understand or access the original data.

Rothenberg had faith that emulation would become the ideal option that archives could take in some near future. However, emulation merely moves the problem of technological catch-

[135]Yeo, "'Nothing is the same as something else,'" 111.
[136]Rothenberg, "Ensuring the Longevity of Digital Information," 2.
[137]Rothenberg, "Ensuring the Longevity of Digital Information," 2..
[138]Rothenberg, "Ensuring the Longevity of Digital Information," 15-16.

up from the record being updated to the programs handling the record needing to be updated, the emulator. John Bennet argued twenty years ago that,

> At the present day, software emulation dominates, being programmable even at the chip level. In an archive, it may be necessary to handle some emulations, but this can only be tenable in the short term, while both the emulated and the host emulator are current in technology terms. Obsolescence for the host environment will bring double jeopardy for the emulated environment. Archiving of an emulation and its dependants should be considered only for the near term, and in the advent of destructive forces.[139]

Emulation systems and the records will suffer the same obsolescence. If an emulator becomes outdated, and no longer works, a new emulator will need to be made or found. Or emulation can be applied again using an emulator to run the emulator. This is not an ideal situation as it adds complexity and creates another point of failure along the workflow. Rothenberg speculated that, "future computers will be orders of magnitude more powerful than ours, future users should be able to ask their computers to generate emulators for obsolete systems on demand."[140] He is assuming a few things here, but mainly that computers will be able to auto-adapt to obsolescence. We don't know what the technological landscape will look like in the future. In the twenty years since Rothenberg was writing, computers have become much more complex and powerful. We can surmise that it is possible for an AI to be programmed to evaluate and correct errors in an emulation system. This is a potential outcome, but it is not yet a reality. The inadequacy of technology at the time of his writing led him to be pessimistic about the implementation of emulation, seeing it as a task for his grandchildren, and wishing them luck.[141]

Other approaches were attempted to bridge the gap between older or foreign systems and current environments. One such was pioneered by Raymond A. Lorie who suggested a method

---

[139]John C. Bennet, *A Framework of Data Types and Formats, and Issues Affecting the Long Term Preservation of Digital Material*, JISC/NPO Studies on the Preservation of Electronic Materials, (London: British Library Research and Innovation Centre, 1997), 20.
[140]Rothenberg, "Ensuring the Longevity of Digital Information," 15.
[141]Rothenberg, "Ensuring the Longevity of Digital Information," 17.

called a Universal Virtual Computer (UVC), which would be able to run software on systems for which the programs were not designed, acting as a universal intermediary.[142] This approach was implemented as a trial at the *Koninklijke Bibliotheek* and National Archives of the Netherlands in the early 2000s. The UVC approach was intended to work on any machine, interpret data objects, and display a reconstructed view of that digital object. This involved a mixture of both emulation and migration, mimicking software and hardware environments and actively migrating files for access.[143] This strategy however was not widely adopted by the digital preservation community. David S. H. Rosenthal, of the LOCKSS project, [144] notes that the UVC approach "was of interest only for preservation, and was thus unable to leverage the emulators that were concurrently being developed by hardware developers, game enthusiasts, and others."[145] Emulators that mimic hardware environments existed at that time, though they were not as well developed. What's more, the UVC approach was only tested on image files, and on a limited number of formats: JPEG, GIF86, and PNG's. The technology seems to have been able to render the structure and meaning: thus it represents a form of active migration using a meta-emulator. It is doubtful that this would have been of any use beyond file level digital objects though, as this functionality represents a more complex problem than simple decoding RGB schemas. Further efforts have been made along these lines, and the results of this progress will be showcased in the next chapter.

Emulation has remained the most recognisable alternative, but Rothenberg's propositions have come under criticism from the archival world. David Bearman offers that,

---

[142]*Raymond A. Lorie, "Long Term Preservation of Digital Information," Proceedings of the 1st ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '01) Roanoke, (2001), 346–352.*

[143]J. Van Der Hoeven, R. Van Diessen, and K. Van Der Meer, "Development of a Universal Virtual Computer (UVC) for long-term preservation of digital objects," *Journal Of Information Science, 31*(3) (2005), 198.

[144]"Dr. David S. H. Rosenthal," LOCKSS, accessed July 5, 2019, https://www.lockss.org/contact-us/dshr/.

[145]David S. H. Rosenthal, "Emulation & Virtualization as Preservation Strategies," *LOCKSS Program, Stanford University Libraries,* (2015), 2.

> Rothenberg is fundamentally trying to preserve the wrong thing by preserving information systems functionality rather than records. As a consequence, the emulation solution would not preserve electronic records as evidence even if it could be made to work and is serious overkill for most electronic documents where preserving evidence is not a requirement.[146]

I believe this represents a point of disagreement in archival thought. Bearman seems to believe that the shown content of a record (the text, image, or sounds) represent the more valuable part of the records' evidentiary nature. Guttenbrunner et al., in looking at video games, sought to preserve their look and feel, with the implication that the records were to be interacted with. This interaction was the point of the games that they were endeavouring to keep accessible.[147] A player does not necessarily care about the resultant content of their time directing Pac-Man: it is the interaction that is important. The interactive look and feel of those games is part of their evidentiary nature, as I argued in the last chapter. This same logic can be extended from games to "playable" documentary forms such as spreadsheets, databases and, in some cases, word processor documents. I will not deny that there is value in the shown content of any record. To Bearman though, the means of interaction, reading, and the copper and transistors of a computer, represent only a means to an end.

Rothenberg's position lies on another side of this discussion, for him the method of interacting with the content has value in and of itself and in relation to the rest of the record and related records. He states,

> If "reading" a document means simply being able to extract its content—whether or not it is in its original form—then we may be able to avoid running the original software that created the document. But content can be subtle: translating from one word processing format to another, for example, often displaces headings or captions or eliminates them entirely. Is this merely a loss of structure, or does it impinge on content as well? If we

[146]David Bearman, "Reality and Chimeras in the Preservation of Electronic Records," *D-Lib Magazine* vol. 5, no. 4 (April 1999), accessed July 5, 2019, http://www.dlib.org/dlib/april99/bearman/04bearman.html.
[147]Guttenbrunner, et al., "Keeping the Game Alive," 65-66.

> transform a spreadsheet into a table, thereby deleting the formulas that relate the cells of the table to each other, have we retained its content?[148]

In my first chapter I made a similar assertion arguing that the interaction is in itself part of the record. When you take away the interactive elements you remove part of the record. Rothenberg takes an extreme position on this asserting that, "any solution to digital preservation that is limited to text will therefore quickly become obsolete. A true long-term solution should be completely neutral to the form and content of the digital material it preserves."[149] He advocates that all digital content should be emulated without regard to its type or scope. Stewart Granger, though a proponent of emulation for archival purposes, criticizes Rothenberg's approach to it claiming that, "Rothenberg's attitude seems to be that if we can't have everything, we can't have anything.[150] Though he is a proponent of emulation Granger does not see emulation as, "a simple, universally applicable, one-time fix."[151] This would involve emulating every digital record in a collection. While I am optimistic that this would not be an impossible task, given current technology it would be a heroic one, as Rothenberg might have put it.

Nonetheless, I do agree with Granger to some extent. Rothenberg's insistence on applying emulation to everything is interesting and worthy of serious thought, but it may not even be desirable. When I defined what I meant by archival interactivity, I intentionally defined it with regard to its level of interactivity. The costs and heroic efforts involved in maintaining the interactivity of gigabytes, petabytes, (and probably more) of data would serve to render the

---

[148]Rothenberg, "Ensuring the Longevity of Digital Information," 10.

[149]Jeff Rothenberg, *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation,*Council on Library & Information Resources (1999), 4.2.

[150]Stewart Granger, "Emulation as a Digital Preservation Strategy," *D-Lib Magazine*, vol. 6, no. 10 (October 2000), Rothenberg's Views of Other Approaches, accessed July 5, 2019, http://www.dlib.org/dlib/october00/granger/10granger.html#ref12

[151]Granger, "Emulation as a Digital Preservation Strategy," Three Positions on Emulation. Granger is quoting Bearman here, who also thought a "magic bullet soloution" was dubious: Bearman, "Reality and Chimeras in the Preservation of Electronic Records."

record even more unwieldy than it already is. Each item in the collection would have to be processed as it is unreasonable to assume that they will all be of the same computer environments and file types.[152] Though I agree with Rothenberg that the loss of structure impinges on the content itself, I feel it is reasonable to find a middle ground between these two sides of this particular archival debate.

Thus I propose that what is saved, in term of interactivity, would become the decision of the archivist in charge of appraisal, and would in fact be treated as an appraisal decision, part of the appraisal process. The archivist would be able to look at the digital record and determine which digital objects were cool enough, that is to say required enough participation or user input, to warrant saving that particular digital object's interactivity. This would slate that record for emulation and the others to a more orthodox (as it presently exists) digital preservation regimen. The methods of emulation will be covered in the next chapter, and which method should reflect the needs of the record and the mandate of the archive. I cannot recommend at what point an archivist would choose to say what constitutes a proper lever of relative coolness to necessitate maintaining the interactivity of the record, as I am of the belief that appraisal decisions such as this rely heavily on circumstances surrounding the idiosyncrasies of that particular record. Nonetheless obvious candidates include spreadsheet files, examples of tech such as anything demonstrating a GUI dependant process such as a game process, operating system, or any prototype program. Any instance where there are separate versions should be likely candidates such as, incarnations of databases, games, webpages, and other programs. Being able to interact with the various incarnations of one of these would give clues as to the thought processes

[152] It is likely that this processing will need to be made automatic. A.I. based automatic ingest and description is a reality. However, the nature and implementation of such systems is beyond the scope of this work.

involved in its development but also provide clues about the workings of the codebase that a researcher could work back from.

An archivist with more Rothenbergian leanings may decide that they should be saving born digital text files and PowerPoint files in addition to these, deeming them cool enough. As these are easier to create they may be more numerous in the record and the number of files may prove prohibitive to emulate the entire collection. This archivist could choose to go ahead and heroically maintain a majority or all of their records' interactivity, or they could choose a sample, a practice with much precedent in the archival world. A survey could be done to determine the nature of the files and a sample could be taken at random intervals from the various discrete units. These units could be series, distinct types of forms or records, or formats depending on the nature of the collection. A random sample could be made of an entire fonds, or series, if the records are homogenous enough. Also files of particular interest could be selected. Whatever the method, as long as they are marked as such a sampling of the collection being emulated may prove enough to get a sense of the interactivity of the remaining files should they be of the same type.

This would have the benefit of keeping the costs associated with emulation down. However, at the present moment, the biggest obstacle is funding, and alternatives such as emulation are very cost prohibitive and require no small amount of expertise. This is echoed in the scholarship. Digital preservation scholar Mike Kastellec posited that finances are the most constraining factor in digital preservation. He saw emulation as an unmastered field that "postpones the processing of data until it is later accessed, potentially allowing greater ingest of information."[153] In this he is completely correct. There are archives that have successfully

[153]Mike Kastellec, "Practical limits to the scope of digital preservation," *Information Technology and Libraries,* 31.2 (2012), 64.

implemented emulation and the Internet Archive can be considered a success. There are new approaches being developed such as the Olive and EaaSi projects, which will be discussed in more depth in the Chapter Three. However the initial investment of time and money towards emulation drives many away. This is compounded by the unknown potential costs involved in maintaining a system of emulation that has not been implemented with visible results. To many, emulation would seem like an optional expense for many archives. Few would be willing to act as a guinea pig for a new system, it being safer to wait till another has gone through the testing and trial phase of a new system.

I take this middle path as I believe that media and format migration still have a part to play in the archival world. Migration does offer a viable alternative for certain aspects of preservation. The Open Archival Information System model (OAIS), for example, provides a workflow that packages information specifically for access, and for long-term preservation.[154] Ideally the long term preservation copy, the Archival Information Package (AIP), would be a bit level copy of the original data. Unless a user requires the bit level data object for one reason or another, an access copy, served as part of the Dissemination Information Package (DIP), acts as the object encountered by a researcher. As often only the rendered content is what is needed, a migrated copy may be good enough. Someone making use of the record for quick reference, to double check a citation or quote for example, may be better served by a migrated copy, not needing to continually use the emulated copy, or having to look at the underlying code.

Migration would often be desirable as original files can often be large and cause a strain on access systems, particularly for video files and newer programs. Migrated files formatted for access tend to be lossey, or have undergone data compression. This refers mainly to the changing

---

[154]The Consultive Committee for Space Data Systems, *Reference Model for an Open Archival System (OAIS),* (Washington DC, June 2012), 2.2.3, 2-7 to 2-8.

of the files from archival formats, such as .PDF for a text file, or .WAV, for an audio file. These are lossless formats, meaning that they do no undergo any compression, and the entire bit level data set is present in the file. These files are good for storage as there is no data loss, but are by nature larger. For many access files, the rendered representation is what is important. They are files that emphasize content over anything else. Thus lossy formats such as .Jpeg, .AVI, or .MP3 are good enough as they are smaller, though lesser in quality. It is worth noting that if a file enters the archive in a compressed format, that file will still be normalized to a lossless format. This can be done by using tools such as Archivmatica, Xena, MIXED, or OpenDataForge. These normalization processes are automatic in that the work is done by the program, but the workflow is fully customizable at almost every level. Some consider these to be unreliable, especially with video formats. Archivist Tyler McNally opted for an ad hoc approach using specialized open source software such as Audacity, HandBrake, Preview, and ImageMagic to process the various files.[155] Whatever method of preservation and access processing the archivist opts for, format, order, and destination are all up to the archivist.

With regard to a digital object's metadata however, migration still poses a risk. Format migration, in changing the bit level data, has the potential to damage or destroy any metadata present in the record. Metadata refers to a part of the digital object that contains information about that particular digital object, its relationships to other documents, and its environment. While as above a migrated copy would have its uses, a bit level copy of the original would maintain all the metadata of that original. This metadata would be valuable to archivists in

---

[155]Tyler McNally, "Practical Digital Preservation: In-House Solutions to Digital Preservation for Small Institutions," Bloggers!, last modified January 23, 2017, https://saaers.wordpress.com/2017/01/23/practical-digital-preservation-in-house-solutions-to-digital-preservation-for-small-institutions/; Tyler goes into more detail in his thesis where he discusses a DIY approach in more depth, Tyler McNally, "Practical Digital Solutions: A DIY Approach to Digital Preservation" (MA thesis, University of Manitoba, 2018).

determining what system the record needed. This idea is not new, as there are semantic units for recording environment information in PREMIS. A working group made up of Angela Dappert, Sébastien Peyrard, Carol C. H. Chou and Janet Delve produced a work detailing the implementation of existing PREMIS 2.0 functionalities that would describe environmental metadata. These semantic units had been rarely used and the paper sought to detail the implementation and capture of environmental metadata.[156] These recommendations were integrated into PREMIS 3.0, which calls for "careful documentation of the technical environment associated with an archived digital object can be an essential component of preservation metadata."[157] Each environment will contain valuable metadata. Such metadata is not just an important part of a complete record, but also in being able to discern the specific hardware that the emulator will need to mimic, as different chipsets, or drivers will have different behaviours.

Part Two: Archival Frameworks

Much of this was to say that the proper tool should be used for the proper job. Emulation does have a place in the archival toolbox. Archives however, cannot simply begin to use emulation. It is a complex tool that influences appraisal, preservation, and access of the records. Its use would influence the entire archival workflow. Archival scholar Greg Bak noted that any improvement in digital preservation "cannot happen without the tight integration of archival or recordkeeping perspectives, while the archival work cannot proceed without a solid understanding of the computer and its functionalities."[158] Digital archiving decisions should reflect an understanding of the original technology and the preservation technology, and what

---

[156]Angela Dappert, Sébastien Peyrard, Carol C. H. Cho, and Janet Delve, "Describing and Preserving Digital Object Environments," *New Review of Information Networking,* 18:2 (2013), 106-173.
[157]PREMIS Editorial Committee, *PREMIS Data Dictionary for Preservation Metadata*, Version 3.0 (June 2015), 251.
[158]Bak, "How Soon Is Now?" 309.

each means for the product of the archival process. Emulation exists as an alternative to either print-to-file, as discussed in Chapter One (essentially a form of media migration equivalent to taking a photograph of an oil painting) or format migration. The implementation of it should still take the essential needs of the individual records, the type of technology, its purpose, and cultural significance into account when preparing the record. This means having at least a minimum of technological literacy across the history of the various mediums, but also an awareness of the path that a record has taken to get to the archive.

Thus one of the problems that emulation is facing is organisational.[159] How does interactivity fall within the problem of how links are established between, "processes that are central to the social phenomenon of making and keeping records."[160] Or, where does interactivity lay within the workflow of archives? I have proposed that interactivity is not just the concern of access and preservation, but of appraisal as well. Appraisal has an effect on access, but as I have described the effect, it pertains to the record's history after it has entered the archive. The history of each record however goes back to its creation. A digital object is created and interacted with through its life cycle but then a distinction is made when it is no longer of use and is deemed archival. The record may also lay dormant for years before it is sent to an archive as part of a collection years or decades after its creation. Through all of that time it was subject to degradation and it had no chance of being of use for any purpose. This emphasis on periods of usefulness characterises the life cycle model that is in use in many Canadian archives. Archivist Jay Atherton explains that the life cycle model as on that "is based on the premise that it is

[159]Devan Ray Donaldson and Elizabeth Yakel make a similar point when looking at the results of PREMIS adoption. They found that organisation was a large factor in the adoption or proper implementation of PREMIS. Devan Ray Donaldson, and Elizabeth Yakel, "Secondary adoption of technology standards: The case of PREMIS," *Archival Science* 13, no. 1 (2013), 57-59.
[160]Glenn Dingwall, "Life Cycle and Continuum: A Review of Recordkeeping Models from the Postwar Era," in *Currents of Archival Thinking,* ed. Terry Eastwood, and Heather MacNeil (Santa Barbra: Libraries Unlimited, 2010), 139.

possible to divide the life of a record into eight distinct, separate stages."[161] He describes it as beginning with a records management phase involving:

- creation or receipt of information in the form of records,
- classification of the records or their information in some logical system,
- maintenance and use of the records, and
- their disposition through destruction or transfer to an archives.[162]

After the record has been deemed of no longer use in its function, it enters the archival phase where it is subjected to:

- selection/acquisition of the records by an archives,
- description of the records in inventories, finding aids, and the like,
- preservation of the records or, perhaps, the information in the records, and
- reference and use of the information by researchers and scholars.[163]

There is another school of thought known as the continuum model which does not seek to delineate the use of the record to specific times of its existence. This model would allow those handling the records to be able to interact with the record's creators more closely and streamline the processes of integration into long term digital preservation and access schemas. Achival scholars following the continuum model assert that a record has use as an object of research or reference from inception. For example a project working with multiple versions of a program may wish to store copies of each version in order to go back and look at what they had done previously, even if only the latest version. They would need to have the versions ordered so as to be accessible. They would also want them to be functioning in order to interact with them. This may even require emulation as older versions may rely on outdated hardware and software. Once created a file would be accessible as a record. A system would need to be set up across an entire

[161]Jay Atherton, "From Life Cycle to Continuum: Some Thoughts on the Records Management-Archives Relationship." *Archivaria* 21 (1986), 44.
[162]Atherton, "From Life Cycle to Continuum," 44
[163]Atherton, "From Life Cycle to Continuum," 44

organisation to allow for the automatic capture and access to allow for a record to be "in use" and "archival" concurrently.

Writing on these two models, digital archivist Glenn Dingwall notes that there was a change in the early twentieth century that redefined the role of the archives, from that of the keeping of records, "by an institution as evidence of its action and for its own future use was replaced by the concept of archives being the repository for records no longer of relevance to the creating institution."[164] This change, Atherton noted, was reflected in the way archivists and record managers viewed each other. Records managers viewed archivists as specialists dealing with documents that have exited the administrative context. Archivists tended to view record keeping as an element in archival practice.[165] Though they saw themselves in each other, the landscape of records management and archival work had become separated. By 1956, Schellenberg had defined archives as "Those records of any public or private institution which are adjudged worthy of permanent preservation for reference and research purposes and which have been deposited or have been selected for deposit in an archival institution."[166] Archives were for storage of research materials. Little other thought was given to their use. Appraisal, rather than interest, determined a record's nature and function.

Of this construction Dingwall claims that, "the life cycle model arose from attempts to describe how recordkeeping processes adapted to accommodate the explosion in the volume of institutional records being created."[167] In the postwar period the capability of informational systems to create documentation became overwhelming. Though transport and communication technology were making the world smaller, it was not feasible to centralize all documentation, as

---

[164]Dingwall, "Life Cycle and Continuum," 142.
[165]Atherton, "From Life Cycle to Continuum," 44.
[166]Schellenberg, *Modern Archives*, 16.
[167]Dingwall, "Life Cycle and Continuum," 140.

much of it was still being actively used. Schellenberg and others were imbued with ideas of the separation of records managers and archivists. This dichotomy led to ideas that the records had distinct phases of use. In practice this system works by appraising records for destruction and/or use, or permanent storage in the archive, with different bodies handling the records at different periods of its life, and disposition acting as the boundary line between the two.

Interactivity can be integrated into the life cycle model, but there are more points in the workflow that need to be negotiated that are beyond the control of the archives. Additionally, it would require digital infrastructure at every point in the life cycle. The need for interactivity would have to be communicated to the records creators and the records managers as a necessary aspect of the record. Certain things like format and metadata would need to be kept in tandem with the record's journey through its life cycle. While paper can be kept relatively easily in its context, electronic records change every time they are used, moved or copied. Their context is more fluid.[168] Additionally they are far more fragile. The life cycle model was fashioned around the needs of the physical documents that were being produced in the immediate post war period. These structures do not translate organically for the digital as there is much risk of change or loss of metadata or binary data.

Archivist Terry Cook maintained a skepticism of this model as it placed archivists and records managers into specific, delineated and passive roles of curation and collection. In the life cycle model the keeping of records is not conceptualized as a proactive process, the responsibility of a senior official who was "often making or strongly influencing the actual decisions themselves, and then preserving the record (or evidence) of those decisions," and

[168]Dingwall, "Life Cycle and Continuum," 153.

therefore taking part in the creation and use of the records. The life cycle conceptualized the management of records as separate from their creation.[169] According to Atherton,

> The archivist serves the needs of the scholar, the historian, and posterity, whereas, the records manager serves the needs of business which is usually profit motivated and which is interested only in information that contributes to or protects that profit or the goals of the organization. To put it another way, the records manager is basically a business administrator and the archivist is basically a historian.[170]

In the life cycle model, the archivist and the records manager fall into roles that are delineated by the needs of a record, and the user of that record, at a point in its life, like a person needing a daycare or hospice. But records are not people. Archival scholar Tom Nesmith would argue that the record does not become something else at any point in time. It has a history, but claiming that at three years a record is now something else and the responsibility of someone else, ignores the nature of any record, in favour of bureaucratic form. The archivist is acting with regard to the form not the record.[171] This form is imparting itself onto the history of the record, artificially.

Archivist Brien Brothman has also levied a critique against this model. This artificial delineation placed on the history of the record characterizes a model that forces understandings of linearity and mortality that pertain to humans onto records. Records though, do not follow the same lines as we do; they do not die.[172] He argues that this is a simplification and that the history of a record is not a linear unidimensional path. Rather is it multidimensional, following many paths concurrently. The life cycle model itself imposes upon the record an end, a teleological destination that falsely simplifies much in the use, purpose and future of a record.[173]

---

[169]Cook, "Electronic Records Paper Minds," 409.
[170]Gerald Brown, "The Archivist and the Records Manager: A Records Manager's Viewpoint," *Records Management Quarterly* 5 (January 1971), 21. In Atherton, "From Life Cycle to Continuum," 43.
[171]Tom Nesmith, "Archives from the Bottom Up: Social History and Archival Scholarship," *Archivaría* 14 (Summer 1982), 10-11.
[172] Brien Brothman, "Archives, Life Cycles, and Death Wishes: A Helical Model of Record Formation," *Archivaria* 61 (2006), 237.
[173]Brothman, "Archives, Life Cycles, and Death Wishes," 268-69.

The continuum model imparts a structure onto the record's timeline, but it does so in a way more in line with the realities and needs of the record. For this, it comes out as the best alternative to the life cycle. The continuum model, as developed in Australia by Frank Upward, Barbara Reed, Sue McKemmish and others, merges records management and archives. This allows for the tracking of a records' changes and interactions over time and space, giving that record a richer contextual relationship to other records and to a range of agents including information creators, managers and users. Although these Australian writers offered a fully-formed idea of a records continuum, Canadian archivist Jay Atherton was the first to use the term, in 1985, in the archival literature.[174] Nonetheless, Upward traces the continuum model back to the ideas of Australian archivist Ian Maclean in the 1950s and 1960s. Maclean thought to combine the archival and records keeping professions, which he saw as not merely related but the same.[175] I am inclined to agree with this sentiment, as I see their roles as more than complementary.

The term "continuum" was not used often till the mid 1990's when archivist Frank Upward published his two-part article "Structuring the Records Continuum," which sought to define and properly think through the idea of a records continuum. It tried to address the question "What should be the relation between archivists and record managers?" To Maclean "it seems that they are or should be the same people, all with similar training but dealing according to personal inclination and aptitude with different periods."[176] Records managers, he notes were, "discovering that marginal record areas, which must be cleaned up before continuous record

---

[174]Atherton, "From Life Cycle to Continuum," 48.
[175]Frank Upward, "In Search of the Continuum: Ian Maclean's 'Australian Experience' Essays on Recordkeeping," in *The records continuum: Ian Maclean and the Australian Archives first fifty years, ed.* Sue McKemmish, and Michael Piggott (Clayton: Ancora Press, 1994), 111-112; Ian Maclean, "Australian Experience in Record and Archives Management," *The American Archivist* Vol. 22 No.4 (October 1959), 389.
[176]Maclean, "Australian Experience in Record and Archives Management," 417.

disposition programs can be effective, require analytical treatment that is just as time-consuming (and hence profit-reducing) as the treatment accorded to archives by archivists."[177] They were doing the same job and in the end were encountering and trying to solve the same problems.

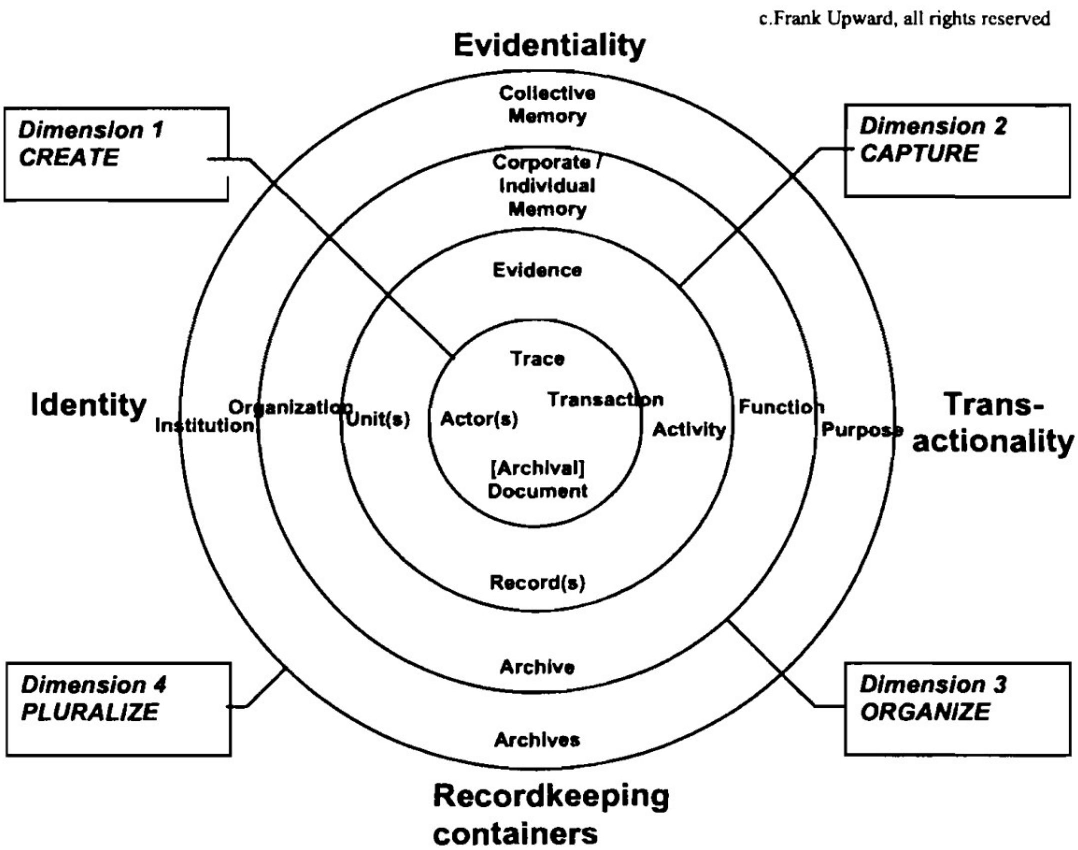


Figure 13: Frank Upward's Records Continuum Model Diagram[178]

This model streamlines the record's timeline by placing that record in a schema where all the aspects of that record's history act upon that record at the same time. In the life cycle model, a record will go through four stages of its life cycle but at no point will it not be subjected to every part of those processes that it has not already gone through. In the continuum model, a record exists simultaneously within all four dimensions of the continuum. It is created; a trace of the record is made. The record is captured, appraised as part of a system. They are organised and

---

[177]Maclean, "Australian Experience in Record and Archives Management," 388.
[178]Frank Upward, "The Records Continuum," in Sue McKemmish, Michael Piggott, Barbara Reed, and Frank Upward, *Archives: Recordkeeping in Society,* (Wagga Wagga: Centre for Information Studies, 2005), 203.

processed into a system of memory, such as an archive. They are then pluralised, made available, and used as per the needs of society for that record.[179] This is represented in Figure 13 by the concentric circles. As the record expands in its scope from the centre to the extremities it is still acting within the other circles. A record is always being added to and so created anew. This can be in the form of metadata, or changed by use and being used for different purposes.

Within these dimensions there are four axes that a record is acting in at all times. The record is never at one point performing only one action. The recordkeeping containers axis represents the context the record inhabits, as a document, its incorporation into a recordkeeping system, its organisation within an archive, which exist as part of societal conceptions of archives. Upward describes the recordkeeping containers axis succinctly as "the objects we create in order to store records."[180] The evidentiality axis characterises how a record exists as a point of evidence of an activity, a process, and a point of memory. The transactionality axis refers to what is happening with the record: it is being created, what its function and purpose is. The identity axis refers to "the who:" specifically who is acting upon the record, which changes depending on how the record is characterized along the other dimensions. This could be the actor creating the records, or the archive that is keeping it as a point of memory. As stated above the record is acting within all eight variables (four axes and four dimensions) at once.[181]

This abstract structure manifests itself right at the creation of the record. In this schema a record is potentially archival from the moment of its inception. There is someone acting upon it, that person is part of an institution, that institution part of a society. As the record exists holistically on many levels, the archivist should see that record with the same view. The record

[179]Frank Upward, "Structuring the Records Continuum - Part One: Postcustodial Principles and Properties," *Archives & Manuscripts*, Vol. 24, Issue 2 (1996), 274.
[180]Frank Upward, "The Records Continuum," 206.
[181]McKemmish et al. *Archives,* 202.

has a function, and is evidence of certain processes and societal contexts. It is then put into a record keeping system where it is at once available for use, for that institution but also as evidence of that institution, in the more traditional archival sense. There are obviously appraisal concerns and privacy and access concerns but those, too, are part of the record. What is important for this analysis is that with regard to access it does not matter if the record is "moments old or a thousand years old."[182]

This is complicated with paper records which can only be in one space at a time. It can be argued that in a digital world the application of this schema is more intuitive. In practice it may manifest as in the following thought experiment set in the near future. An employee working at their desk creates an email and sends it to a co-worker. The email is instantly and automatically logged with regard to its ambient context of the organisation. The email is captured as per the appraisal, including its metadata identifying the sender and recipient, the content, the systems used, the date and time, and a myriad of other factors.[183] The employee may even have the option to put a document in the record for later use, manually, if it does not fit the parameters of automatic appraisal. The email is sent to a server which maintains the record and contains software that manages automated ingest functions, in real time. The server is accessible through any device and requests can be sent for any record.[184] After ingest a request is sent for the DIP, or access copy, that represents the email. The request is logged as part of the history of the email. It is accessed by an employee who needs evidence of that employee's conversation which was

[182]Frank Upward, Sue McKemmish, and Barbra Reed, "Archivists and Changing Social and Information Spaces: A Continuum Approach to Recordkeeping and Archiving in Online Cultures," *Archivaria* 72 (2011), 199.
[183] The Monash research team worked out the specifics of this sort of automated system in their Clever RK Metadata project: Joanne Evans, Sue McKemmish, and Barbra Reed, "Making Metadata Matter: Outcomes From the Clever Recordkeeping Metadata Project," *Archives and Manuscripts 37*(1) (2009), 28.
[184] Barbra Reed goes very deeply into the application of web service technology by recordkeepers that would make such requests and responses possiable. She notes that the technology for this was in wide use in 2010, pg. 128,136, and that implementation of such a system would take at least five to ten years, pg. 136. Barbara Reed, "Service-oriented architectures and recordkeeping," *Records management journal* 20, no. 1 (2010), 124-137.

relevant to an ongoing project. At that same time the email is being accessed by a researcher looking at it as evidence of corporate culture in the early-twenty-first century. Each interaction is logged according to the nature of the request. Ten years later the original employee makes a grave misjudgment and is put under investigation. The email is brought up in an emulation of the original system. The investigators look at the email in the emulation and are able to see the other emails. The emulation itself allows the investigators to see how the employee could have used the program to hide, or manipulate information. They explore the history of the document and see that it was only ever accessed twice, the reasons for that use listed.

This is a very basic and reductive example, but it gives a sense as to the technology involved, and how it could be used within a continuum schema. A digital continuum model could be characterized by a centralised computing system, with decentralised access and work points. One of the benefits that digital objects have with regard to the continuum model is characterised in Frank Upward's complaint:

> what will records keeping and archival processes be like when the location of the material matters less than its accessibility, when the records no longer have to move across clear boundaries in space and time to be seen as part of an archive, and when an understanding that the records exist in space-time, not space and time, is more intuitively grasped by any practitioner.[185]

With networked computer systems it matters less where the server is, or even where the data is, as long as proper preservation is being done and appropriate access protocols are in the control of the archive. Computer infrastructure, such as servers, storage, computers, software and expertise, are all expensive. There are also issues with maintaining adequate security over a record beyond having the data on an un-networked system, which in itself creates problems for access. These factors are prompting more and more repositories to buy cloud-based server space on giants like

[185] Quoted in Dingwall, "Life Cycle and Continuum" 158.

Iron Mountain, Amazon and Google, opting for digital infrastructure as a service.[186] This has its own issues, such as the placement of trust in a corporation whose existence, or structure and mandate, is not guaranteed.[187] Uncertainty aside, as long as a computer is linked to the appropriate network it can serve as an archival work terminal, or as a tool of access. Although legal jurisdictions can complicate record keeping and access, in straightforward functional terms it no longer matters where the record is, as long as the proper preservation protocols are being observed. All other access concerns can be negotiated with interested parties, taking jurisdictional issues into account, with the extreme policy of full, open access over a Web browser as only one option. Numerous limitations and access requirements can be implemented.

Conclusions:

The continuum model is more than just a workflow that will allow for the saving of metadata. It is a way of interacting with records. The continuum model provides a better alternative to the life cycle. The archive can maintain a more detailed history of the record, over the course of a record's existence. It also offers a more streamlined and decentralized access and storage approach that could serve to reduce the strain of implementation of any of the emulation approaches that I will discuss in the following chapter. The closer to creation a record is captured the less work will have to be put into processing it in the long term, thus lowering time to wait

---

[186] "Cloud Data Archiving|Long-Term Object Storage|Amazon Glacier," Amazon, accessed March 31, 2018. https://aws.amazon.com/glacier/; "Archival Cloud Storage – Data Backup and Disaster Recovery| Google Cloud," Google, accessed: March 31, 2018, https://cloud.google.com/storage/archival/; "Iron Cloud Storage," Iron Mountain, accessed: March 31, 2018, http://www.ironmountain.com/digital-transformation/iron-cloud-data-management.

[187] This debate is beyond the scope of this work. Luciana Duranti has done work on this, detailing what an archivist can expect when considering cloud-based internet services. See: Luciana Duranti, and Adam Jansen, "Records in the cloud: authenticity and jurisdiction," in *2013 Digital Heritage International Congress (DigitalHeritage)*, vol. 2, 161-164. IEEE, 2013; Luciana Duranti, "Preservation in the cloud: towards an international framework for a balance of trust and trustworthiness," in *APA/C-DAC International Conference on Digital Preservation and Development of Trusted Digital Repositories*, 5-6. 2014.

for inclusion in the archive. This, along with a more automated ingest system, would allow a record to be accessible sooner. With the capture and processing software distributed throughout an organization's network, expertise that was being used to process and ingest records can be put to use in areas less tedious and better suited to their skills. By distributing workloads away from a processing bottleneck, to a creation-capture approach, the continuum model if implemented well has the possibility to make archival problems more manageable across entire organizations.

I am not suggesting that every archive run completely on a continuum model. Conversion would be costly and time consuming. It would be feasible for archives who have not yet set up an interactive access schema to pay some mind to this organizational method. As with anything setting up an efficient system early will mitigate costs and problems further down the road.

Archives cannot ignore the digital. At present we have been living in a world where digital technologies were normalized into mainstream culture long ago. It has become a redundant activity to advocate for digitalization at this point. The tools and techniques available make it possible to maintain the interactivity of a record. Unfortunately running emulation will be, like much else in the archival profession, akin to the parable of the Red Queen.[188] However as technology progresses the onus lies with archives to maintain an up to date understanding and infrastructure of current systems. These systems are by nature complex, and the systems of access are also changing. As public interaction with information changes, so will that public's expectations. New mediums such as virtual reality environments will change the way people make records. Archives and those that support them by developing tools and techniques will have to keep up. It is unknown what the shape of future mediums will take. Rosenthal posits that

---

[188] "Now, *here*, you see, it takes all the running you can do, to keep in the same place" Lewis Carroll, "Through the Looking-Glass" in *The complete Works of Lewis Carroll* (London: Nonsuch Press, 1966), 152, 150-153. This allusion has struck me as apt. It at once reminds one that while they are dealing with the past the present is still happening. One will never get ahead of the trends in society, but one can try to keep up.

digital access will become more of a specialised field. Where the functioning of the hardware necessary to, "access emulations of legacy digital artefacts will become a specialist item. It will probably be available in library "reading rooms" and academics' offices, but the general public will expect to access these artefacts using a smart-phone."[189] Whatever the reality, techniques of providing proper access to that record should follow the form that was inherent to that record.

What this means is that archivists should seriously consider the application of emulation as a tool that they use and become familiar with. It cannot be implemented haphazardly. Proper and well thought out integration into the archivist's workflow will allow for more effective preservation and access. Archivists should have in their minds a clear idea of what will be appraised for emulation and what not. Inclusion into the archival mandate, and discussion and negotiation with donors, will aide the archive in making these appraisal decisions. This places the responsibility on individual archives to mold a schema that will most effectively bring the record to any interested parties. It is clear though that this and other measures pertaining to digital records have reached a point where archivists can act on them and can start considering different methods. The next chapter will provide an overview of some of the options available to archivist, in order to give a sense of what is possible.

---

[189]Rosenthal, "Emulation & Virtualization as Preservation Strategies," 25.

Chapter Three
It's Messy Behind the Curtain!: Practical Emulation Methods and Technologies

In the previous chapter I made a case for why emulation should be a standard tool of archivists for maintaining the interactivity of digital objects, and how it would fit into current archival methods. This chapter will look at current ways of preserving the interactivity of digital records through emulation. To that end, this chapter will focus on two case-studies, and a number of prospective technologies which represent different technological and organizational philosophies or approaches to the problem of maintaining the interactivity of records.

The first case study concerns the Salman Rushdie papers at Emory University's Manuscripts, Archives and Rare Books Library (MARBL), which offers an example of a successful ad hoc implementation of emulation.[190] This example represents a "do-it-yourself" approach used on a small scale, but is also a good example of the factors acting on an emulated record in an archival setting. My second case study will look at a successful example of large-scale implementation being done by the Internet Archive which makes use of the full variety of open source emulation to provide interactive copies of websites, legacy operating systems, games and other programs.[191]

I will then look at a relatively new approach built off of multiple previous attempts at archival emulation, known as Emulation as a Service. Two groups are currently at the forefront of development, the EaaSI, and Olive emulation projects.[192] This approach simplifies emulation by shifting development, access, and implementation to a third party specializing in emulation of

[190]Laura Carroll, Erika Farr, Peter Hornsby, and Ben Ranker, "A Comprehensive Approach to Born-Digital Archives," *Archivaria,* no. 72 (2011), 61-92.
[191]"Internet Archive," Internet Archive, accessed July 5, 2019, https://archive.org/.
[192]"About EaaSI," Software Preservation Network, accessed June 20, 2019, http://softwarepn.webmasters21.com/eaasi/; "Olive Executable Archive," Olive Archive, accessed: July 5, 2019 https://olivearchive.org.

digital records. This can benefit archives that lack digital infrastructure or staff with technical competencies. This chapter will discuss the strengths and weaknesses of each of these approaches, in order to show their idiosyncrasies as well as the risks and benefits inherent to each approach.

Part One: The Emory project

It should be obvious by this point that implementation of emulation is no simple matter. Despite its complexity archives have achieved positive results in the use of emulation in maintaining the interactivity of digital records. In 2006 the team at MARBL acquired and made available an interactive copy of some of author Salman Rushdie's born digital files. The Salman Rushdie papers were donated to Emory University, who received, in Rushdie's words, "two hundred or so cracked-up falling apart crappy cardboard boxes with stuff just thrown into it without any sense of organisation."[193] The MARBL team appraised and processed that into, "over one hundred linear feet of traditional archival material, such as journals, correspondence, and manuscript writings." [194] They also received many other collections that included some fugitive media, such as floppy disks, CDs, and DVDs, and for the first time the library acquired entire computer systems.[195] In that group of dusty boxes "there were these computers from the stone age of computing, through the iron age, bronze age and so on, there's this bunch of dead computers which had to be exhumed and brought back to life,"[196] these old computers included four Macintosh personal computers, a portable hard drive, and the born digital records that were contained in Rushdie's devices. [197]

[193]Salman Rushdie,"Salman Rushdie on the Opening of His Archive at Emory," YouTube, 4:13, "Emory University," Mar 9, 2010, https://www.youtube.com/watch?v=nqvQKiV-XWA.
[194]Laura Carroll et al. "A Comprehensive Approach to Born-Digital Archives," 63.
[195]Laura Carroll et al. "A Comprehensive Approach to Born-Digital Archives," 63.
[196]Rushdie,"Salman Rushdie on the Opening of His Archive at Emory,"
[197]Rushdie,"Salman Rushdie on the Opening of His Archive at Emory;" "Emory Finding Aids: Salman Rushdie papers, 1947-2008: Series 11, Born Digital Materials," Emory Libraries & Information Technology, accessed June 20, 2019, https://findingaids.library.emory.edu/documents/rushdie1000/series11/?keywords=rushdie.

In appraising Rushdie's collection, the team was forced to deal with the question that every archive must deal with when digital records come in as part of a collection: what is to be done with the hardware and digital records? As I explored in the last chapter, the answer to dealing with digital records is not as simple as taking a paper record and putting it in a folder. The computers were important; Rushdie had spent a great deal of his working time on those computers. Over the course of his career, Rushdie had kept all the computers that he had used in his work, other than his first Macintosh. He had gravitated towards computers for their ability to transmit data instantly, and more securely than mail, a feature that allowed for greater security in the face of threats of violence from radical groups.[198]

Seeing the connection that these systems had to his work, the MARBL team opted for a dual approach of migration and emulation. They created an emulated environment that mimics Rushdie's Performa 5400/180 personal computer, a searchable database of the content in PDF format, and a finding aid for both of these.[199] Geoffrey Yeo noted that properties of the digital make such dual approaches possible, allowing the records to be available to multiple user communities.[200] The project thus used both format migration and emulation. The emulation was built in a way so as to maintain the context and content of the record. Carrol et al. cite Rothenberg's insistence for their decision. "The meaning of a document may be quite fragile, since meaning is in the eye of the beholder: what may be a trivial transformation to a casual reader may be a disastrous loss to a scholar, historian, or lawyer."[201] In discussing the difference between copies of records, one does not know who will be accessing the material or why.

---

[198]Laura Carroll et al. "A Comprehensive Approach to Born-Digital Archives," 63.
[199]Laura Carroll et al. "A Comprehensive Approach to Born-Digital Archives," 80.
[200]Geoffrey Yeo, "'Nothing is the same as something else': significant properties and notions of identity and originality," *Archival Science* 10 104
[201]Laura Carroll et al. "A Comprehensive Approach to Born-Digital Archives," 78; Jeff Rothenberg, "Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation," Council on Library and Information Resources (1999), 4.

This is a good example of the argument that I have made in the previous two chapters. Emulation and the amount of interactivity that you want to save is an appraisal decision. Their decision was based on the reasons made above. The archives had received a "nearly complete record of Rushdie's digital life." This was not a few leftover files; they had a snapshot of the digital life of one of the most consequential and controversial authors of the twentieth century. Thus the MARBL team saw that these computing environments were necessary to understand the nuances of Salman Rushdie's work.[202]

Access was also discussed upon appraisal. "Even in early conversations about researcher access, Rushdie expressed that he did not want the born-digital material to be openly accessible via the Web."[203] The negotiated decision was that access was to be limited to one access point: "all three of these access tools –the searchable database, the emulation and the finding aid – inhabit a single computer workstation in the reading room along with ancillary help documents."[204] The response by extremists to Rushdie's work reinforced the normal concerns regarding the privacy of Rushdie, his family and anyone he may have corresponded with. The result of this is that for a researcher to see the collection they must actually go to the archive.

The decision to do this was made at appraisal. With the idea of how the collection was to be preserved and accessed they were able to form a plan of how to deal with the collection. This decision would not have been made lightly as it increased their workload substantially. As noted above, they opted for a dual approach that utilised both emulation and migration and began by making copies of the digital files. It was agreed however, that the team would not try to recover deleted files. They created a disc image of each hard drive, and "once the hard drives were

[202]Laura Carroll et al. "A Comprehensive Approach to Born-Digital Archives," 64.
[203]Laura Carroll et al. "A Comprehensive Approach to Born-Digital Archives," 68.
[204]Laura Carroll et al. "A Comprehensive Approach to Born-Digital Archives," 80.

duplicated, Peter [Hornsby, a software engineer on the project] began to harvest the metadata for the various types of user generated content. As well as the systems and application files in order to get a more complete picture of the nature and scope of the born-digital material."[205] The team recorded the checksums of each file, creating a baseline copy of the bit level data for later comparison. With this they created a searchable database of the data and metadata.[206] The data consisted of fax, Eudora email messages, and word processing files, which were migrated so that there would be access copies and the original data would not be tampered with.[207]

The MARBL team outsourced the design and functionality of an access interface to Resonance, a web development company. Resonance designed a browser based interface system that would run on a terminal in the reading room. Using this, a researcher could search the database of migrated material, as well as explore an emulation of Rushdie's Performa 5400. The disk image of Rushdie's computer relied on existing emulation software known as SheepShaver. SheepShaver was developed in 1998, and was designed to emulate "MacOS run-time environments" on BeOS and Linux systems.[208] Its status as open source and its capability to work with the Linux based Fedora OS, which the access terminal would use, was ideal. Nonetheless "the software required some configuring by Peter in order to serve the needs of our project, but his communication with developers and hobbyists provided all the information he needed."[209] The emulation was not an exact copy of Rushdie's computer. The team did not merely plug in the hard drives, and boot up what was on them. The uploaded data had been processed into categories of "as is," appearing in both database and the emulation, and "virtual

[205]Laura Carroll et al. "A Comprehensive Approach to Born-Digital Archives," 84.
[206]Laura Carroll et al. "A Comprehensive Approach to Born-Digital Archives," 81.
[207]Laura Carroll et al. "A Comprehensive Approach to Born-Digital Archives," 81.
[208]"SheepShaver: An Open Source PowerMac Emulator," Christian Bauer, accessed July 5, 2019, https://sheepshaver.cebix.net/.
[209]Laura Carroll et al. "A Comprehensive Approach to Born-Digital Archives," 84.

only," files only present in the emulation, into the emulated environment.[210] It was thought better to go about "stripping the emulation and re-populating it with approved data in order to ensure that no restricted data would remain; this also guaranteed only approved content would make it into the MARBL reading room."[211] This meant that the emulation was not a 1 to 1 copy. However the files approved by appraisal to be made available for access were accessible in the emulated original environment with no change in their functionalities.

The MARBL project was made more feasible by the circumstance of only having to set up one instance of emulation on one computer. Because of the access restrictions they only had to troubleshoot for one environment. There would be no worry about needing to adapt the emulation to various browsers and OS environments, which could cause problems with any proprietary software. Though successful at the scale that the MARBL team had decided upon, an increase in user access points, or in the size of the emulated collections, would compound the work. Each system has its own idiosyncrasies that would render any catch-all emulation engine almost impossible, as was discussed with reference to the UVC in Chapter Two. Further the emulation of entire systems is also not a scalable practice, as the processing of entire systems is not practicable, based on the time and power needed to process the amount of data on any system created since the early 1990's in a realistic amount of time.

What the MARBL team did accomplish is a successful adoption of emulation into an archival workflow. They set out to provide a good representation of Rushdie's work without compromising the integrity of the records. Their work shows how emulation, like any other archival tool, can be incorporated into appraisal. Also of note was the positive effect that negotiation with the donor, careful preparation and predetermined goals, had on the

[210]Laura Carroll et al. "A Comprehensive Approach to Born-Digital Archives," 74.
[211]Laura Carroll et al. "A Comprehensive Approach to Born-Digital Archives," 84.

implementation of emulation. This ad hoc approach works best if there is a smaller digital component to the collection, and a lesser degree of variety in the computing environments. This will not be the case forever, though as computer usage increases there will be more and more digital material coming into archives of greater volumes and varieties. MARBL emulated one computer, but it is likely that an archive of a similar size may receive, multiple, mobile phones, PCs, Laptops, hard drives, and tablets all in one collection, all with valuable records worthy of emulation.

Part Two: Internet Archive's Ad Hoc Approach

Adapting existing emulation tools in an ad hoc way is not a very viable solution for most archives. It has however been done successfully, though not in a strictly archival context. The Internet Archive currently maintains millions of pieces of software in browser-based emulations in its software library. The Software Library maintains over 423,000 examples of software that can be interacted with via emulation on the Internet Archives webpage, at the time of writing.[212] This includes everything from games to operating systems, to shareware. One can find individual programs, such as a game like Mortal Kombat, and play it in-browser.[213]

More than just a way to play old games, the main goal of the Internet Archive, when it was started in 1996, was to collect "the public materials on the Internet to construct a digital library," in order to allay a loss of so much digital history. It has since become much more than that.[214] While the Wayback Machine has saved over 356 billion web pages at the time of writing,

[212]"Software Library," Internet Archive, accessed July 5, 2019, https://archive.org/details/softwarelibrary.
[213]"Mortal Kombat by Acclaim/Arena Entertainment/Midway/Probe," Internet Archive, accessed July 5, 2019, https://archive.org/details/gg_Mortal_Kombat_1993Acclaim_Arena_Entertainment_Midway_Probe.
[214]Brewster Kahle, "Archiving the Internet," *Scientific American,*(March 1997), Internet Archive Wayback Machine, accessed July 5, 2019, http://web.archive.org/web/19971011050140/http://www.archive.org/sciam_article.html; Using the Wayback Machine to describe the purpose of the Internet Archive is a great example of this point; Brewster Kahle, and Ana Parejo Vadillo, "The Internet Archive: An Interview with Brewster Kahle," *19: Interdisciplinary Studies in the Long Nineteenth Century,* 21 (2012), accessed July 5, 2019, https://www.19.bbk.ac.uk/articles/10.16995/ntn.760/.

these are not emulations as the majority of these sites work with the current web architecture, their interactivity maintained simply by running them in a browser able to read older scripts.

This chapter is more concerned with the Software Library. On the Library page one can find many examples of emulated systems. A good example is the Windows 3.11 Stock Installation.[215] Internet Archive leveraged open source emulation and compiled an access copy with description on a web browser. This is a clean installation of the Windows 3.11 OS. As this version of windows runs in a DOS shell, it is emulated using an in-browser application of the DOSBox emulator. This emulator is a third party open source emulator built by hobbyists to play old DOS based games.[216] As DOSBox is open source the Internet Archive was able to easily adapt it to their website, for use with any emulations of DOS based programs.

The instance of Windows 3.11 that they emulate is described using the Wikipedia description. The use of the associated Wikipedia article would be problematic with an archival record, as the article is subject to change and may not be complete, or properly cited or vetted. The implementation of the program however is what is important in this analysis. Additionally the access page shows the collections the OS is a part of, and who uploaded it. It also maintains a record of how many people accessed the emulation, and maintains a user presence in the form of reviews so that users can log problems, or give experiential anecdotes of their interaction with the implementation.

The way the emulation is presented is important because most emulations do not run as smoothly as Windows 3.11. Glitches can occur in any emulation: text will fail to format, the

---

[215]"Windows 3.11 Stock Installation," Internet Archive, accessed July 5, 2019, https://archive.org/details/win3_stock.

[216]Discussions with the creators of DOSBox: "Interview with Qbix - 30 April 2008," RBG Classic Games, accessed July 5, 2019, https://www.classicdosgames.com/interviews/peterveenstra.html; "DOSBox 0.73, Interview with the Developers," Sir Authur's Den, accessed July 5, 2019, http://kingofgng.com/eng/2009/06/10/dosbox-073-interview-with-the-developers/; DOSBox Homepage: "DOSBox, an x86 Emulator with DOS," DOSBox, last updated June 26, 2019, https://www.dosbox.com/.

sound will not work, or the graphics will act slowly or dazzle the user with a rainbow of colourful static. These problems are being constantly addressed with the help of user input. For many games, however, the difference in the control setup changes the user's input, and can make a game awkward to play. For example the game Mortal Kombat originally was played with a controller or at an arcade cabinet, but must be played with a keyboard and mouse in the emulation.[217] This type of loss of functionality is unavoidable for many of the titles in the Internet Archive, particularity in their Console Living Room, a collection of games and console emulations.[218] The reality is that most of these systems maintained their own interactive schema, which forces a work around, or a remapping of controls that can make the interaction awkward, at least at first. As with any game learning the controls is part of the experience.

The Internet Archive is an example of a model that could be adapted to smaller archives and be used to maintain access to smaller to medium sized collections on a budget. Though the amount of emulations is still in the tens of thousands, the use of open source emulators for access copies of disc images using in-browser emulation or a specialized access-terminal-based virtual machine would theoretically allow for any archive to do this. Though this sounds relatively easy and thrifty in theory, in practice adapting free software tools to archival records and putting them in a browser, allowing free access, is anything but easy.

The Internet Archive has two advantages that allow this method to work for them. The first advantage is one of purpose. Rather than acting like an archive, the Internet Archive's Software Library is, as its name implies, a library rather than an archive. Like a library, the programs and environments that the Internet Archive maintains are examples of programs that were in use rather than ones that were the result of the actions of a creator in the course of their

[217] Internet archive, "Mortal Kombat."
[218] "Console Living Room," Internet Archive, accessed July 5, 2019, https://archive.org/details/consolelivingroom.

actions. They are generic copies of certain programs, which exist without context to each other, or any provenancial creators. The objects are there to be interacted with not as evidence, but as a reference.

The computers received from donors by an archive will not be the clean copies found on the Internet Archive. They will be contextually linked to the rest of the records, and most files will maintain idiosyncratic changes or differences of the content creator's interactions with the media. That is to say that a computer arriving is unlikely to have the factory state and settings maintained. They will have added files, deleted programs, and options and preferences which will have been changed to suit the original user's interaction with that system. A database developer may have used phpMyAdmin to build a database, but their instance of it will differ greatly from the clean version.[219] The tables and processes set up will conform to their idiosyncratic interactions with the program that they were building the databases in. The organization of those databases in that program would be a valuable tool to emulate as a distinct record. Thus the first advantage of the Internet Archive is one of simplified scope: they do not have to deal with the disorganized and challenging additions and changes that have been made to any computer the writer or reader of this has encountered.

The second advantage that the Internet Archive maintains is that of expertise. Not counting volunteers, the Internet Archive employs over eighty people. Most of these are specialists in software, web-development, digital preservation, and computer engineering.[220] It is possible to adapt open source emulators into access sites, but unless the amount of digital content is small, the number of digital objects being adapted for emulation is low, have a dedicated

[219] I only use phpMyAdmin as an example as I have experience with it.
[220] "Bios," Internet Archive, accessed July 5, 2019, https://archive.org/about/bios.php.

digital preservation and access staff, or, for whatever reason, maintain a large, highly diverse, and specialized workforce, this approach is not generally feasible.

Part Three: Web Content and Rhizome

There are a number of more specialised approaches that seek to capture and display specific media. I briefly discussed the Internet Archive's Wayback Machine earlier, as a good example of this. Like their Software Library, the Wayback Machine is a product of its creator's intentions. It functions as an open repository of saved websites. The sites are organised by URL, and the Wayback Machine's crawlers are continually scanning and recording iterations of the sites over time arranging them in a calendar or a timeline.[221] This is done with the use of passive web crawlers which travel through the web and save as much functionality of a site as possible. The crawler then sends the data to the Wayback Machine, which uploads the content into the timeline of that URL. One of the strengths of this system is that it was started early. There are over 312 billion webpages saved since 1996 and more are added every day. The Wayback machine acted fast and saved a lot of the early web.[222] There are, however, two main drawbacks to this method. The first is that the crawlers only save the internal functionality of a site. The viewer will not be able to interact with any links outside of the URL, imbedded content such as a YouTube video, or flash animations. This leaves much of the content and functionality broken, and any external content, like ads or embedded content, is simply not present, its existence only indicated by an X. Additionally any content located behind a firewall, or any password controlled areas, are blocked from the crawlers. This represents a large portion of the web and

[221]"Wayback Machine," Internet Archive, accessed July 5, 2019, http://archive.org/web/.
[222]A search engine from Oct. 22, 1996 provides a good example of the temporal scope, but captures from this early period are sparser, partly because web pages updated less often. "Lycos, Inc. Home Page," Internet Archive Wayback Machine, accessed July 5, 2019, https://web.archive.org/web/19961022175214/Http://www.lycos.com.

leaves a large blind spot in the record. The intention however was not to present discrete records, but a large picture, and it does act well as a repository of web architecture and design over time.

Another approach to preserving digital records and digital culture is that of the Webrecorder by Rhizome, which focuses on preserving online digital art. Rhizome is a non-profit organisation, which began in 1996 with the goal to find a way of maintaining an online community for digital artists.[223] In 1999 they launched ArtBase as a way of creating a database of online works of art.[224] As online works of art have a high degree of interactivity, it was necessary to use a program that could save the functionality as well as the content of the artworks. Rhizome settled on a platform that used the Baden-Wuerttemberg Functional Long-term Archiving and Access (BwFLA) infrastructure as a base. This has given them the ability to let people interact with the art, as well as storing it. With the full release of Webrecorder software in 2016, Rhizome has since branched into providing an open source online platform for recording web functionality and content.[225]

Rhizome's Webrecorder is a browser-based, user-directed program, which works much like the automatic crawlers used by Google and the Wayback Machine. Rather than an attempt to record the entire internet, it is a way for individuals to record discrete online interactions and content.[226] After logging in to Webrecorder, the user simply types in the URL of where they want to begin, and the recorder records all of the content on each page as the user moves through

---

[223]Matthiew Mirapaul, "ARTS ONLINE; Digital Artworks That Play Against Expectations," *The New York Times,* September 30, 2002, https://www.nytimes.com/2002/09/30/arts/arts-online-digital-artworks-that-play-against-expectations.html.
[224]"ArtBase," Electronic Arts Intermix, accessed: July 6, 2019, http://www.eai.org/resourceguide/exhibition/computer/rhizome/artbase.html.
[225] "Rhizome Releases First Public Version of Webrecorder," Rhizome, accessed July 6, 2019, https://rhizome.org/editorial/2016/aug/09/rhizome-releases-first-public-version-of-webrecorder/
[226]Dragan Espenschied, "Rhizome Releases First Public Version of Webrecorder," Rhizome, last updated August 9, 2016, https://rhizome.org/editorial/2016/aug/09/rhizome-releases-first-public-version-of-webrecorder/.

it.[227] This includes any embedded ads, or videos. The user must take care to browse each page and view all content, so as to get a more complete reproduction. This can include content located behind password protection: the recorder follows where the user goes and records all. Webrecorder compiles each page into bookmarks that can be accessed in a list. Clicking on one of the bookmarks allows the user to interact with every page and bit of content that was originally recorded in its original functionality, including following the links between pages recorded. Thus the more links on a page are clicked the more that can be accessed in review. This technology was developed for artists to record their interactions and online art, but the implications are vast in the realms of archiving and accountability.

This system is limited by it specialized scope, which makes it only useful for certain archival tasks involving specific media. As such it is not as holistic an archival tool as some of the tools that will be discussed shortly. However, a greater number of people, and organisations, are building and maintaining websites. Webrecorder is a useful tool to capture any site that belongs to a donor. Unfortunately, for it to work the prospective site needs to still be accessible in some fashion. As always, capture of any content subject to copyright would have to be part of the donor agreement. That said, the capability for easy capture using Webrecorder would allow any archivist to quickly replicate any online content that falls under the donor agreement, creating a functional record.

There are likely more specialised tools to be found, and likely more to be made. The archivist should be always open to new tools that become available, or become acquainted with prospective tools that may need testing or professional support. There are many areas that can be streamlined, automated, or enhanced with the idea that they could turn into a valuable tool.

---

[227]"Webrecorder," Rhizome, accessed July 6, 2019, https://webrecorder.io/.

Part Four: Emulation as a Service (EaaS)

Although the Emory and Internet Archive examples show both large and small scale implementations of emulation with the goal of preserving the interactivity of the record present, they represent methods that many archives would understandably be pessimistic about. There is, however, a new approach to emulation in which specialised third parties providd emulation as a service (EaaS). Work on EaaS systems has been gaining momentum over the last decade in light of research and development in network and internet technology. Two instances of EaaS are now seeing positive results in testing and implementation. These are the aforementioned EaaSI and Olive projects.

In the early 2010s technologists, seeing the need to provide interactive instances of records, began to look at this challenge in a different manner. Dirk von Suchodoletz, Klaus Rechert and Isgandar Valizada, of the University of Freiburg wrote,

> The emergence of cloud offerings re-centralizes services, and end users interact with them remotely through standardized (web-)client applications on their various devices. This offers the chance to use partially the same concepts and methods to access obsolete computer environments and allows for more sustainable business processes. In order to provide a large variety of user-friendly remote emulation services, especially in combination with authentic performance and user experience, a distributed system model and architecture is required, suitable to run as a cloud service, allowing for the specialization both of memory institutions and third party service providers.[228]

Put succinctly, using server-side computing and storage a specialised organisation can provide emulation in a browser environment to anyone. With this idea the team at Freiburg began the BwFLA project, but they were not the only ones who had thought of this. BwFLA was very similar to the work being done by another collaborative team at Carnegie Mellon University (CMU). The basic premise revolves around the service provider maintaining a server that would

[228]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* vol. 8 issue 1, (2013), 131.

provide the environments needed for any digital records. The user, when interacting with the records, would encounter the record in a browser, imbedded in its original environment.

These technologies do not simplify emulation, but they do move the complexity away from the archivist. This frees more time that can be used for appraisal, preservation, arrangement, and access of the collection, rather than troubleshooting small bits of code. With third party companies specialising in EaaS and support infrastructure, they are able to handle the collections of more than just one archive. This increases their scalability and affordability, potentially making emulation a reasonable, understandable, and predictable expense, instead of a large burdensome expenditure. These technologies, though far along in their development, are, at the time of writing, still not completely ready.

Olive

The Olive Executable Archive is being built at CMU in collaboration with IBM Research. The name Olive is an acronym for "Open Library of Images for Virtualized Execution." It began in 2009 as a way to "jointly explore new ways to work with VIRTUAL MACHINE images" for the purpose of preserving software, as a way of maintaining access to a record of old scientific data.[229] Much scientific, economic, and other evidentiary data and software are born digital. In a paper the CMU team wrote that "at the heart of the scientific method is the ability to reproduce previously-reported results."[230] Experiments need to be done more than once in order to verify results. Having data from just one experiment done once is not enough to do a proper analysis of the observed results, as one point of data does not necessarily confirm or disprove a hypothesis. The digital data, and the software tools used, are needed to

[229]"Origins of Olive," Olive Archive, accessed July 6, 2019. https://olivearchive.org/origins/.
[230]Mahadev Satyanarayanan, Gloriana St. Clairz, Benjamin Gilbert, Yoshihisa Abe, Jan Harkes, Dan Ryanz, Erika Linkez, and Keith Webster, "One-Click Time Travel," School of Computer Science Carnegie Mellon University (June 2015), 1.

reproduce the results, and to compare the original findings to any new results. This is impossible, if the underlying data is locked away in obsolete or otherwise inaccessible databases and applications.

In order to confirm the data, then, we need to have the data and be able to access it. The CMU team cites a controversy concerning the underlying data of an analysis of economic growth published by Reinhart and Rogoff in 2010.[231] Their data was found to be faulty by Herndon et al., who found an error in the calculations in the original MSExcel spreadsheet.[232] The error in the data was discovered due to the reproducibility of the analysis made possible from the interaction with the record. This example raised further questions, one of which is "What if Microsoft goes out of business ten years after the original publication of results, and the Windows environment (which is needed to run Excel) ceases to be in use?"[233] Thus, there is a need have working copies of digital records available, not just the bit level data or the code, in order to test scientific and economic analyses, conclusions and hypothoses. This was the rationale behind the project. Olive works as a stack of layered software that is installed on the archive's systems, which work together "to encapsulate and deliver a bit-exact, pre-packaged execution environment over the Internet."[234] The user would only see the results of these layered systems. The outputs would appear as in figure 1 below, which show screenshots from two successful implementations of the Olive stack. The Olive system would be installed on a standard desktop, or on a server run by the archive, endowed with one or more *x*86

---

[231]Carmen M. Reinhart, and Kennith S. Rogoff, "Growth in a Time of Debt," *American Economic Review*, American Economic Association, vol. 100(2) (January 2010), 573-578.
[232]Thomas Herndon, Michael Ash, and Robert Pollin, "Does High Public Debt Stifle Economic Growth? A Critique of Reinhart and Rogoff," Working Paper 322, Political Economy Research Institute, University of Massachusetts Amherst, (April 2013).
[233]Satyanarayanan, et al., "One-Click Time Travel," 1.
[234]Satyanarayanan, et al., "One-Click Time Travel," 2.

microprocessors, to aid in processing. On that computer would be installed an instance of the Linux operating system, which forms the second layer in Olive's stack of technology.[235]

Figure 14: Windows 3.11, and Oregon Trail emulated using Olive[236]

At a higher level of the stack is a program written at CMU called VMNetX, short for Virtual Machine Network Execution.[237] This program manages the various environments that the Olive system can execute. This program is interesting "in that it allows virtual machines to be stored on a central server and then executed on demand by a remote system."[238] The virtual machines that VMNetX runs are instances of old computing environments that correspond to the digital records. This means that "your computer doesn't need to download the virtual machine's entire disk and memory state from the server before running that VIRTUAL MACHINE. Instead, the information stored on disk and in memory is retrieved in chunks as needed by the next layer up: the virtual-machine monitor (also called a hypervisor), which can keep several virtual machines going at once."[239] The result of this is that the Olive stack can run multiple virtual machines at a time from a single server, allowing multiple users at various access

---

[235]Mahadev Satyanarayanan, "Carnegie Mellon is Saving Old Software from Oblivion," *IEEE Spectrum*, September 19, 2018, https://spectrum.ieee.org/computing/software/carnegie-mellon-is-saving-old-software-from-oblivion.
[236]Satyanarayanan, "Carnegie Mellon is Saving Old Software from Oblivion."
[237]Satyanarayanan, "Carnegie Mellon is Saving Old Software from Oblivion."
[238]Satyanarayanan, "Carnegie Mellon is Saving Old Software from Oblivion."
[239]Satyanarayanan, "Carnegie Mellon is Saving Old Software from Oblivion."

terminals or online via a browser, concurrently, with just a click. When the user clicks a link to an emulation, VMNetX obtains the metadata describing the desired emulation from Olive's web server, and sets up the proper emulator using a program called FUSE.[240] One of the virtual machine's purposes is to run the emulator, which mimics the environment of the needed obsolete software and hardware environment. The emulator runs a copy of the needed operating system, into which is loaded any archival software the records being accessed need to run. Memory is set aside for this in the form of a virtual disk with the storage capacity of the original computer, allowing the computer to act as it would have if one was using the original machine.

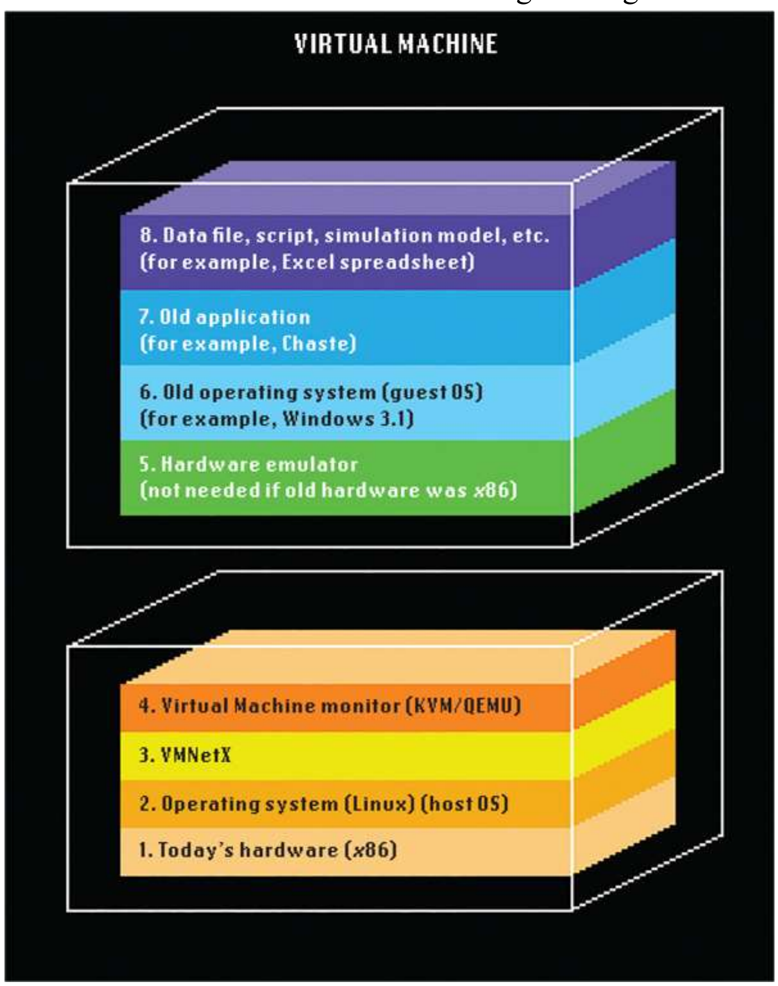


Figure 15: The Olive stack, these abstract layers represent the various software elements that Olive used to open create an environment for digital objects[241]

---

[240] Rosenthal, "Emulation & Virtualization as Preservation Strategies," 8.
[241]Satyanarayanan, "Carnegie Mellon is Saving Old Software from Oblivion."

As of fall 2018, “Olive consists of 17 different virtual machines that can run a variety of operating systems and applications.”[242] They have instances of games such as Oregon Trail (see fig. 1), old websites, and scientific programs such as Chaste 3.1, “short for Cancer, Heart and Soft Tissue Environment, shown below in Figure 16 emulated in Olive. The latter being a simulation package developed at the University of Oxford for computationally demanding problems in biology and physiology”[243] Chaste, though obsolete, is associated with research whose data is still relevant to existing papers and research. The Olive recreation of Chaste features an example of the original data that related to a 2013 paper.

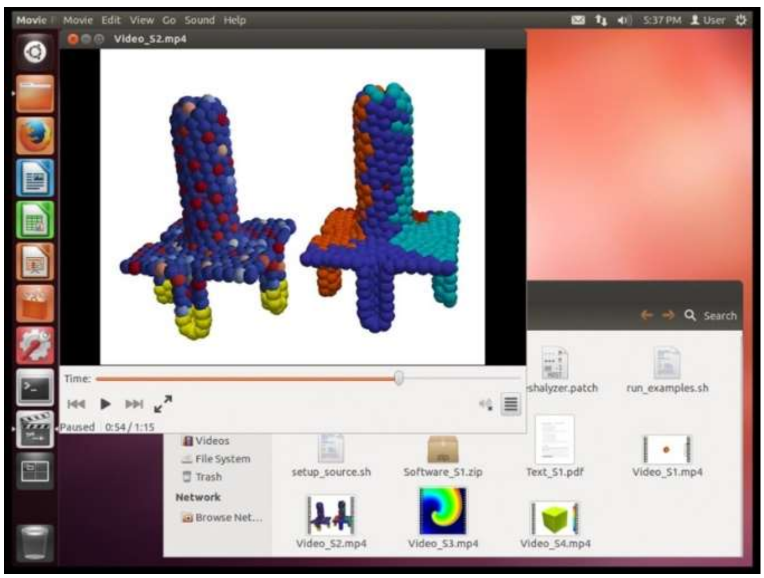


Figure 16: Chaste 3.1[244]

Olive covers a lot of ground with these 17 environments, as most computers use a relatively small number of standardised computing environments. However there are many environments that were less popular, or more unique examples designed for specialised purposes that may fall outside of the Olive virtual machine library. What is more, this library will continually need to be updated as time progresses and new operating systems, software and other

[242] Satyanarayanan, “Carnegie Mellon is Saving Old Software from Oblivion.”
[243] Satyanarayanan, “Carnegie Mellon is Saving Old Software from Oblivion.”
[244] Satyanarayanan, “Carnegie Mellon is Saving Old Software from Oblivion.”

factors evolve and become available. As Olive is a service this would likely be part of an ongoing agreement along with troubleshooting and software patches, and would likely be handled by the company and not individual archives. The Olive system has been tested, but there are still many issues to be resolved before its release.

For the archivist, Olive running on the archive's systems means that initial costs would be tied to processing power and memory needs, as well as the subscription to Olive. While Olive can run on almost any system, the amount of processing power and memory that will need to be available to the application is directly proportional to the number, size, and complexity of the digital records to be emulated. The costs would also be tied to access. An access strategy such as that seen at MARBL, where access is limited to a single terminal in the reading room, would allow for a smaller instance of the system, and reduce set up and infrastructure needs and costs. This could also enable access restrictions based on donor agreements or legal concerns, to be identified during acquisition or appraisal. Alternately, Olive could also enable an approach more akin to that of the Internet Archive's – that is, an open online access system. This would require more output and would place Olive on a server rather than on a single computer. This comes with its own problems, a large one being the security of a networked server. While the MARBL approach allows for more control, the Internet Archive approach allows for greater reach. The answer for many archivists may lie in between these two. Every collection will present different challenges to be decided upon appraisal.

EaaSI: It's Messy Behind the Curtain

The other EaaS project in development is the Scaling Emulation and Software Preservation Infrastructure (EaaSI) project being run out of Yale. It is much more ambitious than Olive, but EaaSI inherits the work of its long pedigree made up of previous attempts at creating a viable system for archival emulation. There are seeds of the EaaSI system in Jeff Rothenberg's idea of an umbrella system to handle all the possible variations of environment. He hypothesised about such a system in the late 90s when he asserted, "an ideal approach should provide a single, extensible, long-term solution that can be designed once and for all and applied uniformly, automatically, and in synchrony (for example, at every future refresh cycle) to all types of documents and all media, with minimal human intervention."[245] This assertion of a single and once and for all approach was taken seriously and the UVC project, seen in the previous chapter, attempted to make this a reality. The UVC ultimately failed, but many of the lessons learned from the endeavour were transferred into the next generation of emulation projects.[246]

The BwFLA project succeeded these efforts. The project was headed by a team working at the University of Freiburg who were focused on the preservation of digital artworks.[247] The goal would not be limited to art, however. They sought to create a system that would provide provide Emulation as a Service (EaaS) to archives that would be more scalable than more ad hoc approaches. The project created a working proof of concept that would have in theory hosted none of the content but provided emulation functionality from a cloud server that the archive would connect to. EaaS would provide the environment and the archive would provide the record and the access point.

The BwFLA site boasts a working example of Vilèm Flusser's hypertext model work and lectures from 1989. This demo emulates the environment of the Mac Performa 630, even adding textures to mimic the use of a CRT monitor.[248] Within that environment, Flusser's program can be interacted with and his lectures viewed.[249] The user can also interact with the rest of the

---

[245]Rothenberg, Avoiding Technological Quicksand," 16.
[246]Rosenthal, "Emulation & Virtualization as Preservation Strategies," 2.
[247]Klaus Rechert, Isgandar Valizada, Dirk von Suchodoletz, and Johann Latocha, "bwFLA – A Functional Approach to Digital Preservation," *Praxis der Informationsverarbeitung und Kommunikation* (2012), 259.
[248]Rosenthal, "Emulation & Virtualization as Preservation Strategies," 7-8.
[249]"The Digital Heritage of Vilèm Flusser," bwFLA, accessed July 6, 2019, http://eaas.uni-freiburg.de/demo-flusser.html.

system. The service is presentation only, which means that the user is only interacting with that iteration of emulation that appears on the screen; they are not directly interacting with the referenced access copy. Any changes they make in that system are non-existent once that representation is closed. Access to the system is controlled by the archive, and BwFLA offered compatibility with privacy and authentication software like Shibboleth to aid in tracking access and maintaining security.[250]

The ideas behind the EaaS model that BwFLA developed are being applied and elaborated on at Yale University with funding from the Andrew Mellon Foundation, the Alfred P. Sloan Foundation, as well as partnerships with the University of California, Stanford University, Notre Dame University, the University of Toronto, Bibliothèque Nationale de France, and Carnegie Mellon University.[251] These teams benefitted from being able to build on existing technology. "When we began work on the EaaSI project, all of these features were in nascent form thanks to the hard work of the OpenSLX, the team of developers spun-out from the BwFLA team at the University of Freiburg to provide commercial support and development for the EaaS product."[252] The new product is a lot more robust than the previous iterations of EaaS as the primary goal of EaaSI is to make EaaS scalable to any collection.[253] This would mitigate any tinkering on the part of the archives, and make large scale emulation in an archival environment a realistic alternative.

---

[250]"Home – Shoboleth Concepts – Shibboleth Wiki," Shibboleth, accessed July 6, 2019, https://wiki.shibboleth.net/confluence/display/CONCEPT/Home; bwFLA, "DPA 2014 - 'Research and Innovation' finalists: Presentation by bwFLA Team," Vimeo, 26:35, Digital Preservation Coalition, 2015, https://vimeo.com/106569098.
[251]"EaaSI Participants," Software Preservation Network, accessed July 6, 2019, https://www.softwarepreservationnetwork.org/EaaSI-participants/.
[252]Euan Cochrane, "Making Things EaaSIer: Overview from EaaSI's PI," Software Preservation Network, accessed July 6, 2019, https://www.softwarepreservationnetwork.org/blog/making-things-EaaSIer-overview-from-EaaSIs-pi/.
[253]Klaus Rechert, "EaaSI Software Development Update," Software Preservation Network, accessed July 6, 2019, https://www.softwarepreservationnetwork.org/blog/beta-rollout-development/.

EaaSI acts as an interconnected system of shared emulation, at the center of which is a centralized node, which acts as a repository of pre-configured software environments. The number of these environments will need to correspond to the variety of computing technology. This goes far beyond the 17 environments that Olive offers. To make this a reality, “the EaaSI program of work is committed to configuring at least 3000 emulated or virtualized configured computing environments during the first grant-funded phase of work.”[254] Of the 3000 environments most will be variations of the more common OS’s. This will not cover the myriad of variation that will be needed for the archival context, which contains an almost unlimited variety of hardware, OS, and program suites. As a way to leverage the distribution network, the EaaSI team added a feature that allows an institution acting as a node host to share any ready-made software environments.

This network-sourcing permits one archive to share a software environment with the whole network. This sharing helps to avoid any replication of labour so that any archive can instantly benefit from work already done without having to invest in their own work or creating an ad hoc system.[255] This also allows for the sharing of expertise in a system that may be unknown, and to facilitate conversations concerning practical preservation problems that may not have happened had this relationship between archives not been present. This repository, which can be seen below in Figure 18, can be tailored for the user to some extent and environments can be uploaded and shared, or kept private as per user agreements or in accordance with legal, or security concerns. These environments act as the shells that the digital objects can be loaded into.

An environment is more than just the saved state. Metadata about these environments is needed in order for users to interact with the environments. For this EaaSI again leverages the

---

[254]Rechert, “EaaSI Software Development Update.”
[255]Rechert, “EaaSI Software Development Update.”

network by using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH).[256] OAI-PMH is a tool that harvests metadata from systems over a network. This metadata is then uploaded to a database much like the Environments database, allowing archives to share and download metadata to create a more complete catalogue.

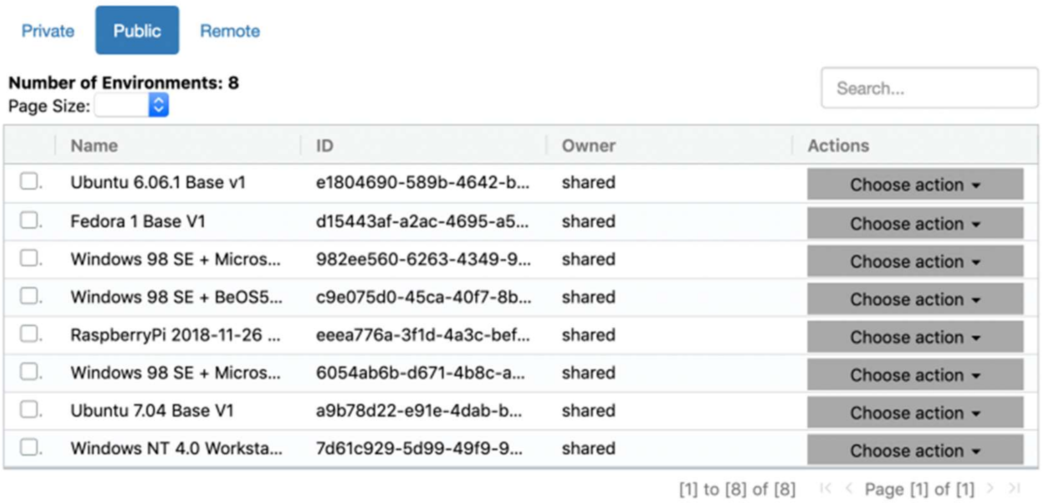


Figure 17: "Screenshot of EaaSI landing page featuring search results list of configured environments in the EaaSI Network"[257]

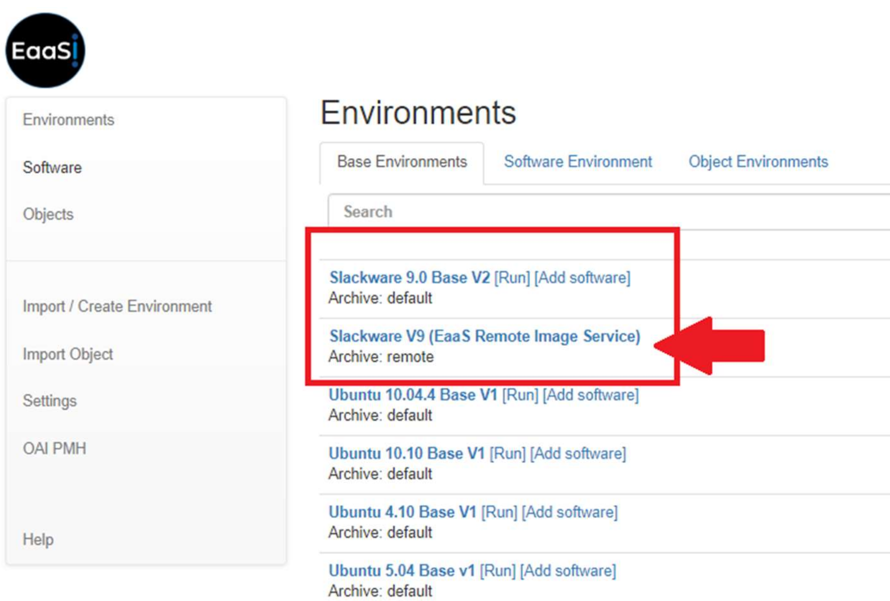


Figure 18: The environment repository user interface showing different operating systems which can be used to make thousands of derivative environments[258]

---

[256]"Open Archives Initiative Protocol for Metadata Harvesting," Open Archives Initiative, accessed July 6, 2019, https://www.openarchives.org/pmh/.
[257]Rechert, "EaaSI Software Development Update."
[258]Cochrane, "Making Things EaaSIer."

These environments are accessed from a browser which cannot run the environment. An emulator is needed in order to connect the access site in the browser to the software environment. As there are numerous environments there will, by necessity, need to be numerous emulators. Thus EaaSI has made a third repository specifically for emulators.[259] The environments, metadata, and emulators uploaded by a node can be set to private or public depending on the need or sensitive nature of the collection.

These centralised repositories act as a library that other parts of EaaSI can use to carry out the emulations. One of the major parts of EaaSI is the *Universal Virtual Interactor (UVI)*. This system is being constructed in order to fulfill one of the goals of the EaaSI project, which is to have multi-file digital objects automatically match to the correct and "best" configured environments. Euan Cochrane, Digital Preservation Manager at Yale University Library, describes this as an "automagic interaction." The goal of getting the correct match is to make sure that the environment and digital object work correctly but also to make the objects appear and work as intended, with as little user input as possible.[260] This would be in sharp contrast with a process where the archivist or user would load a program or virtual device, pick an emulator and input the digital objects, and then hope that there would be few errors or glitches.

The UVI's matching works by having "File format identification tools such as DROID and Siegfried match file format standards to files' structures by looking for patterns within the files that match to patterns the file format requires there to be in the files, those patterns are called 'signatures'."[261] This relies on certain applications having identifiable patterns in the files

[259] Rechert, "EaaSI Software Development Update."
[260] Euan Cochrane, "Designing a Universal Virtual Interactor (UVI) for Digital Objects," Digital Preservation Coalition, last updated November 23, 2018, https://dpconline.org/blog/idpd/designing-a-uvi-for-digital-objects#_ftnref2.
[261] Cochrane, "Designing a Universal Virtual Interactor (UVI) for Digital Objects," A Better Way.

they create, and then identifying those patterns. This also relies on understanding what metadata corresponds to which environment. The UVI's ability to do this relies on the quality of the repository: "the UVI will only ever be as useful as the set of available interaction environments and their metadata." The bigger the repository, the better the matching tools can function.

Once the digital objects are matched to the environment, the environment would be loaded into the emulator and the file(s) loaded into that environment, so that they are available for interaction by the user. This is done by taking the requested file and putting it in an empty image (like a hard drive, but virtual) and then loading that disk image into the environment. This can be done a number of ways, as explained by Cochrane,

1. Upon request, EaaSI opens the disk image of the configured interaction computing environment and puts the file to be interacted with into the "startup" location for operating systems that support that (e.g. variants of Microsoft Windows and Apple Mac OS)
    a. For files that can't fit within the free space of the environment's disk image, EaaSI puts a link in the startup location to the file stored in another drive that was created on-demand with the file within it
    b. This requires the file's mime type to be associated with the appropriate interaction application within the operating system such that the required application is configured as the one to automatically open that mime type, and that information has to be available as metadata for EaaSI to use - all things we are implementing within EaaSI
2. Create a script/program that opens the file using the operating system and required applications' available methods and include it in the operating system's start up process. For example running LibreOffice Writer in Linux with the following bash command should open the file "file.odt"
   libreoffice --writer file.odt
   Then use either the method above, or e.g. in the case of linux operating systems or e.g. MS-DOS, edit the startup sequence to include the appropriate command.
3. Use a pre-recorded GUI macro to automatically open the file using mouse and keyboard maneuvers. This is probably the most widely applicable option but may require restricting manual user-input until the file is open which could have various challenges associated with it (such as identifying when to re-enable user input).[262]

---

[262]Cochrane, "Designing a Universal Virtual Interactor (UVI) for Digital Objects," Automatically Opening the files for Interaction. This may be a lot of detail, but it does show the complexity of the endeavour as well as the realistic.

The process of making the digital objects appear in the emulation properly is not always easy, however as more environments are processed and understood, and metadata collected, the process should become more streamlined and intuitive, with only the obscure and unique environments and objects needing attention.

The idea is ambitious but there is no reason to only work towards a halfway solution. This does however make the UVI very complex. There is still work to be done in making the process fully automatic as well as improving the response times. The UVI has been shown to work, but the difficulty lies in the variety of the environments and digital objects, as well as the need to have it work smoothly and error free most, if not all, of the time.

The archive's end of the process is much simpler than other approaches. EaaSI's software is downloaded on the archive's computer infrastructure which connects the archive's systems to the EaaSI servers. This establishes the archive as a node on the EaaSI network. The environments and emulation software are all held on the main server, and can be accessed though the EaaSI user interface (UI). This saves the archive storage space, processing power, and the necessity to build an original UI, or fit it into an existing or proprietary one that may cause another point of failure for the system. When opening a record in the UI the archive sends the request to the EaaSI server which then reads the metadata indicating the necessary environment. The environment is opened remotely using the OpenSLX technology that grew out of the BwFLA project. The OpenSLX system was designed to remotely implement and manipulate an OS over a network. The system allows not only this, but has the ability to run multiple OS's on multiple systems at the same time.[263] Unlike the Olive stack and the MARBL system, which are stored and run completely on the archive's computers, EaaSI's functionalities, as well as the OS,

[263]"Meet the OpenSLX Portal," OpenSLX, accessed July 6, 2019, https://slxbox.c3sl.ufpr.br/?do=Main.

and environments, are completely offsite, in the cloud. All archival records being opened in EaaSI are held by the archive on their storage systems, or on the server of a data storage service. The archive only needs to install and set-up the connective software.

This is not an all or nothing system. Once implemented, EaaSI offers total control of the records being emulated for the archive. The system was built with archival and digital preservation contexts in mind. Thus with regard to issues of appraisal, the archivist can choose what they want, or can have, available for interactive access down to the item level. The team has also understood that archives cannot share everything everywhere to everybody, because of donor agreements, various pieces of national and subnational legislation such as privacy laws, and ethical issues including professional codes of ethics and Indigenous or other community-determined access protocols. EaaSI looks to offer a virtual reading room that allows the archive to control access. An archive can limit access to a reading room, a network or an organisation, as with Emory's Salman Rushdie collection. Or, like the Internet Archive, they can have it freely available over the internet.

Another feature that is under development is the ability to save the results of any interaction within the environments. This would give a researcher the ability to reload their unfinished interaction as separate files to continue work. The results of these interactions can be reviewed by archival staff for inclusion or disposal as per the archive's practices and judgement. The outputs of the interactions can also interact with devices like printers so that researchers can fully interact with the environments by giving them the ability to use whatever functions the original environment was capable of. These features can be turned on or off based on the restrictions surrounding the records, and the availability of peripheral devices.

The phrase that Mark Suhovecky, the Data Curation Developer at Notre Dame, used to describe the progress in April 2019 characterizes the EaaSI system well: “It's Messy Behind the Curtain.”[264] The system is being designed to be record-focused, rather than process-focused. To borrow McLuhan’s terms, the record may be cool, but the path to it should be as hot as possible. As of writing in May 2019, the system is undergoing Beta testing, which began in February 2019.[265] All the component parts of EaaSI need to be tried, and the network itself must be put to the test. Suhoveky writes that since “Of course there are bugs!”, there will be a great deal of testing on the system. This level of testing though indicates that the system’s framework has been successfully constructed and is now merely being subjected to stress in the hopes of finding points of failure and weakness that could be addressed. The project is scheduled for completion in June 2020.[266]

The system would require some sort of subscription to the service that would correspond to the level of usage required. Like the Olive system, EaaSI has done most of the work and the primary concerns of the archivist would be their normal acquisition, appraisal, and digital preservation duties with the addition of initial setup on the archive’s systems, negotiation over the features needed, as well as fitting the system in to their unique workflow. In the long term, subscription fees, trouble shooting, and maintenance will be the largest concerns. Compared to setting up an entire system, the outlay is minimal.

Once implemented, EaaSI’s concerns will revolve around the recuperation of operating costs as they will need to pay technicians to aid in set up and troubleshooting, as well as server

[264]Mark Suhovecky, and Seth Anderson, “Beta Deployment: EaaSI in AWS or Bust!,” Software Preservation Network, accessed July 6, 2019, https://www.softwarepreservationnetwork.org/blog/beta-deployment-EaaSI-in-aws-or-bust/.
[265]Seth Anderson, “EaaSI Update Nov 2018 – Jan 2019,” Software Preservation Network, accessed July 6, 2019, https://www.softwarepreservationnetwork.org/blog/EaaSI-update-nov-2018-jan-2019/.
[266]Mike Cummings, “Project Revives Old Software, Preserves ‘Born-digital’ Data,” Yale News, last updated February 13, 2018, https://news.yale.edu/2018/02/13/project-revives-old-software-preserves-born-digital-data.

fees for in-cloud processing and storage. In order to limit the need to develop a massive computer infrastructure, the system will be hosted on Amazon Web Service. The distributed nature of the service however does not just allay the problems surrounding the gathering of metadata and environmental data, it also serves to spread the cost. The more archives that subscribe to such a service the lower the overall costs, and with more use the system will become more robust and efficient, after initial concerns would be swept away.

Conclusions:

The technologies and approaches detailed in this chapter represent ways of dealing with a few problems that an archivist will have to tackle. The initial problem was that of reproducibility. Can an original environment be reproduced accurately and repeatedly? Another is scalability: how much can the archive reproduce, how often and to how many? Scale is limited by the size and number of files, the variety of formats or types of data and the complexity of the systems. The reach that the archivist intends for the records can also have an effect.

MARBL, at Emory University, was successful in their ad hoc solution for the Salman Rushdie papers by creating an emulation of the original environment for a limited audience. In the long run it is not feasible for an archive to opt for this type of ad hoc method, unless the scope is small in size and complexity and the access burden requirements are low – or if the archive is gifted with an emulation expert with substantial free time, and a generous budget. The Internet Archive's Software Library is also not feasible for the same reasons. The Internet Archive had the expertise, funds, and will to create such a service, but few other archives will share these necessary conditions.

This does not mean that an archive should shy away from maintaining the interactivity of their digital holdings. As preservation is often time dependant, it may be wise to opt for a short-term solution such as maintaining original hardware, or format migration (provided that bitidentical copies of the originals are kept). These are viable options that demonstrably work now. The Olive and EaaSI projects represent technologies that are being tested and refined now.[267] Should the testing go well they will be more scalable and will be more reproducible than any previous technologies, including the UVC, BwFLA, or any do-it-yourself method. There may be hesitation about outsourcing, but an archivist can and should be as critical and interrogative as they need to be of these parties in the interest of maintaining the integrity of their records.

In the short run, this means getting out of the comfort zone, investigating options, and seeing what is needed for your particular institution and situation. In the long run there is serious work to be done with the backlog of digital content, much of it on portable media or on donors' hard drives and cloud spaces, which have not gone through proper preservation and is currently doomed to a slow death by bit rot. It may be necessary to go through existing collections to reappraise them for digital content – including in this process an appraisal for emulation. Inquiring and engaging in projects like Olive and EaaSI will only bring greater familiarity with the methods, technologies, and possibilities surrounding emulation. Understanding these tools will allow the archivist to see the options available for the collection, appraisal, preservation and access of these records. This will have a cost, but digital preservation and access is a necessary expense.

[267]EaaSI and Olive are still in development at the time of writing, and even if an archivist does decide to use such a service, implementation does not happen overnight. Relying on original systems or format migration may offer a necessary stopgap solution during the wait for these EaaS systems to come online.

Conclusion

An understanding of the present society, culture, and the people who participate in it, requires interaction with our digital legacy. We as individuals are thoroughly integrated into a massive digital infrastructure that we use to augment our inherent capacities for communication, memory, and awareness of ourselves, our communities, and the world. These information networks are more than just a means to an end; they are also massive sites of the creative energies of individuals and groups. This digital infrastructure is ever-changing: an ongoing creation emerging from diverse efforts of many individuals. The numerous pieces of hardware and software that link, for example, our bank accounts to Paypal, to Amazon, to UPS or to Canada Post, were all the creative work of someone (or some team).

When we use these networked and cloud-based services we are not engaging with a static system. Users of these systems change them with their curiosity, their wants, hopes and frustrations, and simply through traffic-driven network effects. The companies that use, maintain and develop these networks write programs, and create better hardware. These innovations sit on the shoulders of creative work done by others who came before. Every new version of software, set of research data, piece of code, newly designed user interface, database, language, game, piece of art, and interface is an act of creativity that affects the lives of everyone. For example by making it easier to navigate a point of sale system and allowing for a less frustrating day at work for a clerk; or by creating a work of art and posting it to YouTube or a forum and giving people a chance to explore ideas or feelings. The effects of our digital labours are subtle and complex, but not incomprehensible; they affect our actions and rely on our expectations. These quotidian efforts of digital creativity are valuable and important.

The archivist should look to preserve these acts of creativity, whether mundane or stimulating, in a way that conveys the greatest nuance that can be conveyed. Allowing the records to be as cool as possible will give the individual a chance to warm it with the heat of their participation. Though archivists may seek to preserve hotter versions of records, when making appraisal decisions relating to digital objects a cooler record may have more value. The interactive elements of certain digital records may be invaluable to those records. This thesis sought to show ways in which preserving this interactivity may be possible, and to provide a rationale for a tentative optimism in making that a reality. The Olive and EaaSI projects are just two of the ways in which an archive can begin to explore the task of providing interactive access to the digital records that have come to be so ubiquitous in our daily lives.

Though I stated that each digital record should be encountered with as much nuance as possible, I am aware of the amount of effort that this statement implies. There is a large variety in types of digital records and, like non-digital records, they must be treated with regard to their properties, which correspond to the purpose or function of that digital object. These distinctions can be broken down further into format and the digital environments the files were designed for. Additionally, the size of digital collections is increasing. Extent, variety as well as complexity combine to make processing digital collections a costly and time consuming task. I have argued that emulation is the most faithful way to access and preserve digital content. It does not destroy, overwrite or replace the original bit level data, as format migration does, and has the potential to allow the user accessing the record to interact with the record in the manner that it was originally designed for.

As there is such a variety in digital objects, and an almost equal variety of digital environments, many have been justifiably pessimistic about emulation's adoption as an archival tool. The new approaches to emulation that were discussed in chapter three, and those like them that may come forth in the future, show more and more promise. It would be fruitful for archivists at every level to think about how emulation fits into their institutions, including how it might interact with their conceptual models already in place (for example, life cycle or continuum), or if it can be integrated in a way that takes the nature of the records into account and streamlines the workflow at their institution. This may involve looking into third party tools, and starting conversations with software developers and with other professionals working in memory institutions worldwide online.

With emulation comes the problem of appraisal. The archivist must decide whether or not the interactive elements of a record are significant to that record. Christoph Becker argues that, since hardware can be replaced with emulation and software can be replaced through format migration, therefore "significance lies in the performance."[268] He notes that any argument between migration and emulation is an argument of means as opposed to ends. I have argued that the preservation of the interactive properties of digital records is an end, such that the performance of that record should be as close to its original performance (look and feel) as possible for access.[269] The means that decisions about whether migration or emulation are employed should align with the distinct nature of the digital record being processed.

The problem of appraisal may not be as complex as getting an old program to work in a virtual machine, but it deserves serious thought. Again as different digital objects have different properties, each must be treated with regard to those properties. I stated that a digital record's interactivity becomes an essential element when a user's input is integral for its use and

---

[268] Christoph Becker, "Metaphors We Work By: Reframing Digital Objects, Significant Properties, and the Design of Digital Preservation Systems," *Archivaria* 85(Spring 2018), 23.
[269]Becker, "Metaphors We Work By," 23.

comprehension, and when a resulting output, culminating in a set of data to be consumed, is not the primary significance of the media.[270] Performance of the record is important, but as digital records vary in function and purpose, the interactive elements may not always be a significant property. This is meant to be a starting point rather than a hard and fast distinction. It forces the archivist to ask the question: how necessary is interactivity to this record?

To explore how one may go about answering this, I drew upon McLuhan's ideas concerning participation with media, and particularly his catchy labels of "hot" and "cool."[271] These are used to indicate the contribution a user must make to engage with the medium. This is a relative gauge as one medium, such as a book, may require little input beyond turning pages while reading and being aware of language, and the imagery that the author uses. A video game such as Pac-Man forces the user to decide a path, making it "cooler" than a book, while a computer game such as Minecraft, which provides tools and a setting while requiring the user to input imagination, story, planning and construction, is even "cooler" than Pac-Man.

This relativity blurs the lines as to what constitutes a need for emulation. It does however provide a heuristic to think through what makes a record interactive and how important that interactivity is to the value of the record, as determined by the archivist in light of provenance, institutional mandates and the wishes of the donor. This is why the appraisal of digital objects requires an understanding of the nature of the objects, including an understanding of what makes a record hot or cool. To do this, archivists must understand digital formats, and basic concepts relating to size and purpose. What's more, the archivist should understand why a particular record or system relies upon a mouse, or a joystick, or a touchpad, and why the hypertextual connections that make up a digital object are important. I am not suggesting that the archivist

[270] Chapter One, Page 5.
[271] McLuhan, *Understanding Media*, 23.

become expert in touch screen or virtual reality technology, but rather to cultivate a reasonable awareness and familiarity with the various technologies, and their historical emergence and significance. This understanding of the nature and history of digital interactivity would allow the archivist to make a more reasoned appraisal decision.

Unfortunately an archivist without a system to emulate their digital holdings would have little inclination to consider a digital record's interactive properties. It has been one of the aims of this thesis to act as a brief overview of the way various approaches to emulation work, such as the ad hoc solution at Emory University, the massive, purpose-built solution engineered by the Internet Archive, and the scalable, networked EaaSI, and Olive projects. This effort was made in order to bring familiarity to those that may not have often encountered the range and variety of emulation solutions currently in use or under development. I have covered the technical aspects of these new approaches in the third chapter. I will conclude by making one last argument in favour of these systems.

When implemented properly, emulation-as-a-service (EaaS) systems such as Olive and EaaSI will make the archive more accessible and interesting. Interesting can also be read: fun. This may seem like a superficial argument, but I believe it is an important one. Archives are unfamiliar places and concepts for most people. Interaction with archival records is often an unavoidable tedium of searching through ledgers and cross-referencing information among correspondence, reports and accounts. Public awareness of the archives is a subject of interest to Tom Nesmith. Speaking on the public perception of archives he noted that, "they can seem remote from our day-to-day concerns and important community issues. They may simply seem to be where old documents go to die and gather dust. Nothing could be farther from the truth."[272] Anyone who has opened a box, or folder, and discovered a new story, perspective, or information, has seen the way that archives can change the way they encounter the world around them. Many archival researchers have these "archive stories" and see the value in bringing a meaningful engagement between the archive and society, culture, and history, despite the inevitable tedium of archival research.[273]

Making digital records available interactively over a widely used software application, such as a Web browser, will not only make records more accessible, but will make them accessible in a more intuitive and interesting manner. Through format migration, one can be given an isolated file to read. By also using emulation, one can be given access to the file in its original environment, and can see it in relation to its environment and other contextually linked records. Users would be able to guide their own interaction with the records as they move through the environment to access the file they want to interact with, rather than downloading an archival PDF and disregarding the now broken interactive paths. The act of access becomes part of the experience. It may still be a series of spreadsheets accessed within an old operating system, but the users would be doing more than merely reading.

The implementation of an emulation system can only increase the potential scale of the archival audience. Nesmith notes that the role of archives has expanded in scope and purpose, to the point where archives "have the potential to be a central force in the creation of new knowledge.[274] Historians are not the only audience for these records. These systems would make

---

[272]Tom Nesmith, "The Missing Piece: Towards New Partnerships with Users of Archives," *CITC Canadian Issues/Themes Canadiens: The Canadian Archives Summit: Towards a New Blueprint for Canada's Recorded Memory* (Special Edition, 2014), 54.
[273]Antoinette Burton edited a volume looking at the ways archives have played a part in various scholar's lives, how it shaped their work and what archives mean to them. Antoinette Burton, ed., *Archive stories: facts, fictions, and the writing of history,* (Durham: Duke University Press, 2006).
[274]Tom Nesmith, "Toward the archival stage in the history of knowledge," *Archivaria* 80 (2015), 122.

any open emulated records available to as many users as can access a terminal or a computer with internet access that the servers can support. Not all records would be open; the Rushdie project in chapter three is a good example of a closed emulation access system. The access rights and restrictions the archive placed upon the records would determine the level and nature of access that may be available to a user. What ever the restrictions or limitations, the adoption of such a system will allow archives everywhere to more effectively preform one of archives' paramount objectives, having people use the records.[275]

To conclude, I posit that it maybe time to turn down McLuhan's thermostat in the archive. I have encouraged archivists to look into these technologies and start to think about providing access to digital records in ways that align with the practices and capabilities of the society that they were made in. The reproducibility and scalability of the EaaS systems I have discussed show promise, but none of the options are foolproof. In the end, there is no "automagic bullet," to borrow a phrase from Euan Cochrane.[276] These systems are still in their infancy and will require much testing, input, and negotiation between the providers and the archival community.

Thus it is important to engage with the emulation specialists, and give them feedback. These systems show promise, but they are not perfect, nor are they beyond criticism. The input that an archivist can give may help these systems to become more robust, or to demonstrate their limits and so help set the foundations for a more effective system.

---

[275] I know that there are opinions as to what the primary objective of an archive is, I am of the opinion that though preservation and organization are fundamental, these are done so that the records can be used by people, for whatever purpose. Though the reasons for this may vary from organization to organization

[276] Euan Cochrane, "Designing a Universal Virtual Interactor (UVI) for Digital Objects," Digital Preservation Coalition, last updated November 23, 2018, https://dpconline.org/blog/idpd/designing-a-uvi-for-digital-objects#_ftnref2.

As with any new system there will be a cost. There will be an outlay of money and labour. How much cost is reasonable should be up to the archivist and archival institution, maximizing their individual efforts through the structured collaboration of open source, cloud-based development. To add to these problems are the frustrations of the task itself. Constant troubleshooting, negotiation, and tech support will be a certainty, as with the implementation of any new system. However, action and interest taken now will save effort later. Having an emulation system integrated into the archives' workflows before a collection of 20 petabytes of data shows up will allow the archive to more easily and cheaply process it and make it available in a timelier manner. The alternative is having to quickly set up an ad hoc and insufficient preservation and access scheme, which leaves problems for future archivists.

A relatively recent example of this can bee seen in the case study of the Organizing Committee for the 2010 Olympic and Paralympic Winter Games (VANOC). The City of Vancouver Archives (CVA) job was to set up acquisition and appraisal procedures at relatively short notice to handle the large amount of records that were coming into the CVA from the 2010 Olympic and Paralympic Winter Games. This included a large amount of digital records that the team was not fully prepared to accept.[277] The existing digital preservation and appraisal infrastructure at CVA was insufficient to the task. New workflows and had to be created to deal with the digital records.[278] Had these problems been engaged with at an earlier time, the CVA team would not have better been able to deal with this "worst-case scenario" of limited time and large diversity of records derived from a large organisation that was about to be shut down.[279]

[277]Courtney C. Mumma, Glen Dingwall, and Sue Bigelow, "A First Look at the Acquisition and Appraisal of the 2010 Olympic and Paralympic Winter Games Fonds: or, SELECT * FROM VANOC_Records AS Archives WHERE Value="true"," *Archivaria* 72 (Fall 2011), 93-122.
[278]Mumma, et al., "A First Look," 120-121.
[279]Mumma, et al., "A First Look," 120.

Courtney Mumma et al. are right to say their case was "representative of the impending deluge of digital records that looms over archivists ill-equipped to deal with it."[280] Their case shows a need to prepare proper procedures, infrastructure, and expertise for any digital records that may show up at the archives' door. At worst archivists unfamiliar with digital preservation may make mistakes as they gain experience. This could result in damage to the records, going over time and budget, or having the final result of a project be poorly organised and accessible. However lessons will be learned as the records are processed. At best, it will also allow for archivists to provide the curious people of the future with a sense of who we are now in a nuanced manner that maintains the integrity of the record.

[280]Mumma, et al., "A First Look," 94.

Bibliography

Allen, Max. "Interview with Ted Nelson." Ideas CBC Radio, 1979. Accessed: July 5, 2019 https://archive.org/details/ideas-maxallen-tednelson.

Amazon. "Cloud Data Archiving|Long-Term Object Storage|Amazon Glacier." Accessed July 5, 2019. https://aws.amazon.com/glacier/.

Andersson, Otto Emanuel. *The Bowed-Harp: A Study in the History of Early Musical Instruments.* New York: AMS Press, 1930.

Anderson, Seth. "EaaSI Update Nov 2018 – Jan 2019." Software Preservation Network. Accessed July 6, 2019. https://www.softwarepreservationnetwork.org/blog/EaaSI-update-nov-2018-jan-2019/.

Atherton, Jay. "From Life Cycle to Continuum: Some Thoughts on the Records Management-Archives Relationship." *Archivaria* 21 (1986): 43-51.

Bak, Greg. "How Soon Is Now? Writings on Digital Archiving in Canada from the 1980s to 2011." *The American Archivist* vol. 79, No. 2 (Fall/Winter 2016): 283-319.

Bak, Greg. "Not Meta Just Data: Redefining Content and Metadata in Archival Theory and Practice." *Journal of Archival Organization*, 13:1-2 (2016): 2-18.

Bardini, Thierry. *Bootstrapping: Douglas Engelbart, Coevolution, and the Origins of Personal Computing*. Stanford: Stanford University Press, 2000.

Barnes, Susan B. "Alan Kay: Transforming the Computer into a Communication Medium," *IEEE Annals of the History of Computing*, vol. 29 issue 2 (April-June 2007): 18-30.

Bauer, Christian. "SheepShaver: An Open Source PowerMac Emulator." Accessed July 5, 2019. https://sheepshaver.cebix.net/.

Bearman, David. "Reality and Chimeras in the Preservation of Electronic Records." *D-Lib Magazine* vol. 5, no. 4 (April 1999). Accessed July 5, 2019. http://www.dlib.org/dlib/april99/bearman/04bearman.html.

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

Bennet, John C. *A Framework of Data Types and Formats, and Issues Affecting the Long Term Preservation of Digital Material*. JISC/NPO Studies on the Preservation of Electronic Materials. London: British Library Research and Innovation Centre, 1997.

Berners-Lee, Tim. "A Brief History of the Web." World Wide Web Consortium (1994). Accessed July 5, 2019. https://www.w3.org/DesignIssues/TimBook-old/History.html.

Berners-Lee, Tim. *Weaving the Web: The Original Design and Ultimate Design of the World Wide Web by its Inventor.* San Francisco: Harper San Francisco, 1999.

Brooker, Katrina. "“I Was Devastated”: Tim Berners-Lee, the Man Who Created the World Wide Web, Has Some Regrets." *Vanity Fair,* accessed: July 3, 2019. https://www.vanityfair.com/news/2018/07/the-man-who-created-the-world-wide-web-has-some-regrets.

Brothman, Brien. "Archives, Life Cycles, and Death Wishes: A Helical Model of Record Formation." *Archivaria* 61 (2006): 235-269.

Brown, Adrian. *Practical Digital Preservation.* Facet, 2014.

Brown, Gerald, "The Archivist and the Records Manager: A Records Manager's Viewpoint." *Records Management Quarterly* 5 (January 1971).

Burton, Antoinette ed. *Archive stories: facts, fictions, and the writing of history.* Durham: Duke University Press, 2006.

Bush, Vannevar. "As We May Think." *The Atlantic,* (July 1945). Accessed July 5, 2019. https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/.

Bush, Vannevar. "As We May Think." *Life Magazine*, vol. 19, no. 11, September 10, 1945: 112-124.

BwFLA. "DPA 2014 - 'Research and Innovation' finalists: Presentation by bwFLA Team." Vimeo, 26:35, Digital Preservation Coalition, 2015. https://vimeo.com/106569098.

BwFLA. "Emulation as a Service – bwFLA." Accessed: July 20, 2018. http://eaas.uni-freiburg.de.

BwFLA. "The Digital Heritage of Vilèm Flusser." Accessed July 6, 2019. http://eaas.uni-freiburg.de/demo-flusser.html.

Campbell-Kelly, Martin, William Aspray, Nathan Ensmenger, and Jeffrey R. Yost. *Computer: A History of the Information Machine*. Boulder: Westview Press, 2013.

Carroll, Laura, Erika Farr, Peter Hornsby, and Ben Ranker. "A Comprehensive Approach to Born-Digital Archives." *Archivaria,* no. 72 (2011): 61-92.

Carroll, Lewis. "Through the Looking-Glass." in *The complete Works of Lewis Carroll.* London: Nonsuch Press, 1966.

Ceruzzi, Paul E. *A History of Modern Computing,* 2nd ed. Cambridge: The MIT Press, 2003.

Cochrane, Euan. “Designing a Universal Virtual Interactor (UVI) for Digital Objects.” Digital Preservation Coalition. Last updated November 23, 2018. https://dpconline.org/blog/idpd/designing-a-uvi-for-digital-objects#_ftnref2.

Cochrane, Euan. “Making Things EaaSIer: Overview from EaaSI’s PI.” Software Preservation Network. Accessed July 6, 2019. https://www.softwarepreservationnetwork.org/blog/making-things-EaaSIer-overview-from-EaaSIs-pi/.

Cook, Terry, “Electronic Records, Paper Minds: The Revolution in Information Management and Archives in the Post-Custodial and Post-Modernist Era,” *Archives & Social Studies: A Journal of Interdisciplinary Research* vol. 1, no. 0 (March 2007): 399-433.

Cook, Terry. “It's 10 o'clock: Do you know where your data are?” *Technology Review* Vol. 98 Issue 1 (1995): 48-53.

Cortada, J.W. "The ENIAC's Influence on Business Computing, 1940s-1950s." *Annals of the History of Computing, IEEE* 28, no. 2 (2006): 26-28.

Cummings, Mike. “Project Revives Old Software, Preserves ‘Born-digital’ Data.” Yale News. Last updated February 13, 2018. https://news.yale.edu/2018/02/13/project-revives-old-software-preserves-born-digital-data.

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

Dickey, Thomas. “Lynx2.8.9.” Last updated July 8, 2018, https://lynx.invisible-island.net/lynx2.8.9/index.html.

Dingwall, Glenn. “Life Cycle and Continuum: A Review of Recordkeeping Models from the Postwar Era.” in *Currents of Archival Thinking,* ed. Terry Eastwood, and Heather MacNeil, 139-162. Santa Barbra: Libraries Unlimited, 2010.

Donaldson, Devan Ray, and Elizabeth Yakel. "Secondary adoption of technology standards: The case of PREMIS." *Archival Science* 13, no. 1 (2013): 55-83.

DOSBox. “DOSBox, an x86 Emulator with DOS.” Last updated June 26, 2019. https://www.dosbox.com/.

Doug Engelbart Institute. “1968 Demo - Douglas C. Engelbart, PhD et.al.” Accessed July 7, 2019. http://dougengelbart.org/pubs/video/fjcc68/Englebart's-1968-Transcription.html.

Duranti, Luciana, and Adam Jansen. "Records in the cloud: authenticity and jurisdiction." In *2013 Digital Heritage International Congress (DigitalHeritage)*, IEEE vol. 2, 161-164. 2013.

Duranti, Luciana. "Preservation in the cloud: towards an international framework for a balance of trust and trustworthiness." in *APA/C-DAC International Conference on Digital Preservation and Development of Trusted Digital Repositories*, 5-6. 2014.

Edwards, Benj. "The Computer Mouse Turns 40." *Macworld* (December 9, 2008). Accessed July 5, 2019, https://www.macworld.com/article/1137400/input-devices/mouse40.html.

Electronic Arts Intermix. "ArtBase." Accessed: July 6, 2019. http://www.eai.org/resourceguide/exhibition/computer/rhizome/artbase.html.

Emory Libraries & Information Technology. "Emory Finding Aids: Salman Rushdie papers, 1947-2008: Series 11, Born Digital Materials." Accessed June 20, 2019. https://findingaids.library.emory.edu/documents/rushdie1000/series11/?keywords=rushdie.

English, William K., Douglas Engelbart, and Melvyn L. Berman. "Display-Selection Techniques for Text Manipulation." *IEEE Transactions on Human Factors in Electronics*, vol. HFE-8, no. 1 (March 1967): 5-15. Accessed July 5, 2019, https://web.stanford.edu/dept/SUL/library/extra4/sloan/MouseSite/Archive/AugmentingHumanIntellect62/Display1967.html.

Engelbart, Christina. "Inventing the Mouse." *Historic, Collective IQ Review,* April 12, 2013. https://collectiveiq.wordpress.com/2013/04/12/inventing-the-mouse/#prettyPhoto.

Engelbart, Douglas. "Augmenting Human Intellect: A Conceptual Framework." SRI Summary Report AFOSR-3223, Prepared for: Director of Information Sciences, Air Force Office of Scientific Research. SRI International, (October 1962). The Douglas Engelbart Institute, accessed July 5, 2019. http://dougengelbart.org/content/view/138.

Engelbart, Douglas C. X-Y Position Indicator for a Display System. US Patent 3541541, filed June 21, 1967, and issued November 17, 1970.

Espenschied, Dragan. "Rhizome Releases First Public Version of Webrecorder." Rhizome. Last updated August 9, 2016. https://rhizome.org/editorial/2016/aug/09/rhizome-releases-first-public-version-of-webrecorder/.

Evans, Joanne, Sue McKemmish, and Barbra Reed. "Making Metadata Matter: Outcomes From the Clever Recordkeeping Metadata Project." *Archives and Manuscripts* vol. 37 issue 1 (2009): 28-56.

Forstrom, Michael. "Managing Electronic Records in Manuscript Collections: A Case Study from the Beinecke Rare Book and Manuscript Library." *American Archivist*, vol. 72 (Fall/Winter 2009): 460–77.

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.

Google. "Archival Cloud Storage – Data Backup and Disaster Recovery| Google Cloud." Accessed July 5, 2019. https://cloud.google.com/storage/archival/.

Granger, Stewart. "Emulation as a Digital Preservation Strategy." *D-Lib Magazine*, vol. 6, no. 10 (October 2000). Accessed July 5, 2019. http://www.dlib.org/dlib/october00/granger/10granger.html#ref12

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*, Issue 1, vol. 5 (2010): 64-90.

Herndon, Thomas, Michael Ash, and Robert Pollin. "Does High Public Debt Stifle Economic Growth? A Critique of Reinhart and Rogoff." Working Paper 322, Political Economy Research Institute, University of Massachusetts Amherst, (April 2013).

Internet Archive. "Bios." Accessed July 5, 2019. https://archive.org/about/bios.php.

Internet Archive. "Console Living Room." Accessed July 5, 2019. https://archive.org/details/consolelivingroom.

Internet Archive. "Lycos, Inc. Home Page." Wayback Machine. Accessed July 5, 2019. https://web.archive.org/web/19961022175214/Http://www.lycos.com.

Internet Archive. "Mortal Kombat by Acclaim/Arena Entertainment/Midway/Probe." Accessed July 5, 2019.https://archive.org/details/gg_Mortal_Kombat_1993Acclaim_Arena_Entertainment_Midway_Probe.

Internet Archive. "Internet Archive." Accessed July 5, 2010. https://archive.org/.

Internet Archive. "Software Library." Accessed July 5, 2019. https://archive.org/details/softwarelibrary.

Internet Archive. "Wayback Machine." Accessed July 5, 2019. http://archive.org/web/.

Internet Archive. "Windows 3.11 Stock Installation." Accessed July 5, 2019. https://archive.org/details/win3_stock.

Iron Mountain. "Iron Cloud Storage." Accessed July 5, 2019. http://www.ironmountain.com/digital-transformation/iron-cloud-data-management.

Isaacson, Walter. *The Innovators: How a Group of Hackers, Geniuses, and Geeks Created the Digital Revolution*. New York: Simon & Schuster, 2014.

Kahle, Brewster. "Archiving the Internet." *Scientific American* (March 1997). Internet Archive, Wayback Machine. Accessed July 5, 2019. http://web.archive.org/web/19971011050140/http://www.archive.org/sciam_article.html.

Kahle, Brewster, and Ana Parejo Vadillo. "The Internet Archive: An Interview with Brewster Kahle." *19: Interdisciplinary Studies in the Long Nineteenth Century,* 21 (2015). Accessed July 5, 2019, https://www.19.bbk.ac.uk/articles/10.16995/ntn.760/.

Kastellec, Mike. "Practical limits to the scope of digital preservation." *Information Technology and Libraries* 31.2 (2012): 63-71.

Kate & Corwen Ancient Music Ancient Instruments. **"**Bowed Lyre ~ Shetland Gue." accessed, July 3, 2019**.** http://www.ancientmusic.co.uk/string.html#gue**.**

Kay, Alan. "User Interface: A Personal View." in *Multimedia: From Wagner to Virtual Reality,* ed. Randall Packer, and Ken Jordan, Expanded ed. 121-131.New York: W. W. Norton & Company, 2002.

Landau, Valerie. "How Douglas Engelbart Invented the Future." *Smithsonian Magazine* (January 2018). Accessed: July 5, 2019. https://www.smithsonianmag.com/innovation/douglas-engelbart-invented-future-180967498/.

Landau, Valerie, and Eileen Clegg. *The Engelbart Hypothesis: Dialogs with Douglas Engelbart.* Next Press, 2009.

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.

Licklider, J.C.R. "Man-Computer Symbiosis." *IRE Transactions on Human Factors in Electronics,* vol-1 (March 1960): 4-11.

LOCKSS. "Dr. David S. H. Rosenthal." accessed July 5, 2019. https://www.lockss.org/contact-us/dshr/.

Lorie, Raymond A."*Long Term Preservation of Digital Information.*" Proceedings of the 1st ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '01) Roanoke, 2001: 346–352.

Maclean, Ian. "Australian Experience in Record and Archives Management." *The American Archivist* Vol. 22 No.4 (October 1959): 387-418.

McKemmish, Sue, Michael Piggott, Barbara Reed, and Frank Upward. *Archives: Recordkeeping in Society.* Wagga Wagga: Centre for Information Studies, 2005.

McLuhan, Marshall. “Marshall McLuhan 1974 – Full Lecture Living in an Acoustic World | University of South Florida.” YouTube, 54:09, mywebcowtube. May 2, 2016. https://www.youtube.com/watch?v=0l_ugK386QY&t=1465s.

McLuhan, Marshall, Quentin Fiore, and Jerome Agel. *The Medium Is the Massage,* New York: Random House Inc., 1967.

McLuhan, Marshall. *Understanding Media: The Extensions of Man*. Cambridge: The MIT Press, 1994.

McNally, Tyler. “Practical Digital Preservation: In-House Solutions to Digital Preservation for Small Institutions.” Bloggers! Last modified January 23, 2017. https://saaers.wordpress.com/2017/01/23/practical-digital-preservation-in-house-solutions-to-digital-preservation-for-small-institutions/

McNally, Tyler. “Practical Digital Solutions: A DIY Approach to Digital Preservation.” MA thesis, University of Manitoba, 2018.

Mirapaul, Matthiew. “ARTS ONLINE; Digital Artworks That Play Against Expectations.” *The New York Times,* September 30, 2002. https://www.nytimes.com/2002/09/30/arts/arts-online-digital-artworks-that-play-against-expectations.html.

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

Nelson, Ted. “Project Xanadu Founded 1960 The Original Hypertext Project.” Accessed July 5, 2019. http://xanadu.com/.

Nelson, Ted. “Ted Nelson Demonstrates XanaduSpace (by Arthur Bullard).” YouTube, 7:31, TheTedNelson, July 10, 2013. https://www.youtube.com/watch?v=1yLNGUeHapA.

Nelson, Ted. “Ted Nelson – Vannevar Symposium 1995.” YouTube, 41:30, Colin McDonnell, July 13, 2015. https://www.youtube.com/watch?v=A3mptA5sZTM.

Nelson,Theodore H. *Computer Lib: You can and must understand computers now / Dream Machines: New freedoms through computer screens—a minority report*. 1974.

Nesmith, Tom. "Archives from the Bottom Up: Social History and Archival Scholarship." *Archivaría* 14 (Summer 1982): 5-26.

Nesmith, Tom. “The Missing Piece: Towards New Partnerships with Users of Archives.” *CITC Canadian Issues/Themes Canadiens: The Canadian Archives Summit: Towards a New Blueprint for Canada’s Recorded Memory* (Special Edition, 2014): 54-58.

Nesmith, Tom. “Seeing Archives: Postmodernism and the Changing Intellectual Place of Archives.” *American Archivist*, Vol. 65 (Spring/Summer 2002).

Nesmith, Tom. "Toward the Archival Stage in the History of Knowledge." *Archivaria* 80 (2015): 119-145.

Olive Archive. “Olive Executable Archive.” Accessed: July 5, 2019. https://olivearchive.org.

Olive Archive. “Origins of Olive.” Accessed July 6, 2019. https://olivearchive.org/origins/.

Open Archives Initiative. “Open Archives Initiative Protocol for Metadata Harvesting.” Accessed July 6, 2019. https://www.openarchives.org/pmh/.

OpenSLX. “Meet the OpenSLX Portal.” Accessed July 6, 2019. https://slxbox.c3sl.ufpr.br/?do=Main.

Personal Computer Museum. “Home.” Accessed July 5, 2019. https://pcmuseum.ca/index.asp.

Polachek, Harry. "Before the ENIAC." *IEEE Annals of the History of Computing* 19, no. 2 (1997): 25-30.

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

RBG Classic Games. “Interview with Qbix - 30 April 2008.” Accessed July 5, 2019. https://www.classicdosgames.com/interviews/peterveenstra.html.

Rechert, Klaus Isgandar Valizada, Dirk von Suchodoletz, and Johann Latocha. “bwFLA – A Functional Approach to Digital Preservation.” *Praxis der Informationsverarbeitung und Kommunikation* (2012): 259–267.

Rechert, Klaus. “EaaSI Software Development Update.” Software Preservation Network. Accessed July 6, 2019. https://www.softwarepreservationnetwork.org/blog/beta-rollout-development/.

Reed, Barbara. "Service-oriented architectures and recordkeeping." *Records management journal* 20, no. 1 (2010): 124-137.

Reinhart, Carmen M., and Kennith S. Rogoff. “Growth in a Time of Debt.” *American Economic Review,* American Economic Association, vol. 100(2) (January 2010): 573-578.

Rekrut, Ala. "Material literacy: reading records as material culture." *Archivaria* 60 (2006): 11-37.

Rhizome. “Webrecorder.” Accessed July 6, 2019. https://webrecorder.io/.

Rosenthal, David SH. "Bit preservation: A solved problem?." *International Journal of Digital Curation* 5, no. 1 (2010): 134-148.

Rosenthal, David SH. "Emulation & Virtualization as Preservation Strategies." *LOCKSS Program, Stanford University Libraries* (2015): 1-37.

Rothenberg, Jeff. *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation.* Council on Library & Information Resources, 1999.

Rothenberg, Jeff. "Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation." Council on Library and Information Resources (1999).

Rothenberg, Jeff. "Ensuring the Longevity of Digital Information." Rev ed. (1999): 1-18.

Rushdie, Salman. "Salman Rushdie on the Opening of His Archive at Emory." YouTube, 4:13. "Emory University," Mar 9, 2010. https://www.youtube.com/watch?v=nqvQKiV-XWA.

Satyanarayanan, Mahadev. "Carnegie Mellon is Saving Old Software from Oblivion." *IEEE Spectrum*, September 19, 2018. https://spectrum.ieee.org/computing/software/carnegie-mellon-is-saving-old-software-from-oblivion.

Satyanarayanan, Mahadev, Gloriana St. Clairz, Benjamin Gilbert, Yoshihisa Abe, Jan Harkes, Dan Ryanz, Erika Linkez, and Keith Webster. "One-Click Time Travel." School of Computer Science Carnegie Mellon University (June 2015): 1-13.

Schellenberg, T.R. *Modern Archives: Principles and Techniques*. Chicago: The Society of American Archivists, 1998.

Shibboleth. "Home – Shoboleth Concepts – Shibboleth Wiki." Accessed July 6, 2019. https://wiki.shibboleth.net/confluence/display/CONCEPT/Home.

Society of American Archivists. "Format Migration." accessed July 5, 2019. https://www2.archivists.org/glossary/terms/f/format-migration.

Society of American Archivists. "Media Migration." accessed July 5, 2019. https://www2.archivists.org/glossary/terms/m/media-migration.

Software Preservation Network. "About EaaSI." Accessed June 20, 2019. http://softwarepn.webmasters21.com/eaasi/.

Smillie, Keith. "Early Computing at the University of Alberta and the Introduction of the LGP-30." *IEEE Annals of the History of Computing,* vol. 29 (2007): 65-73.

Smith, Trevor F. "Memex #001." Accessed: July 5, 2019. https://trevor.smith.name/memex/l.

Smith, Trevor F. “Memex #001 Demo.” YouTube, 5:51, Trevor F. Smith, June 19, 2016. https://www.youtube.com/watch?v=pW4SS_9nXyo.

SRI International. “1968 "Mother of All Demos" with Doug Engelbart & Team (1/3).” YouTube, 34:45, Doug Engelbart Institute, March 12, 2017. https://www.youtube.com/watch?v=M5PgQS3ZBWA.

SRI International. “1968 "Mother of All Demos" with Doug Engelbart & Team (2/3).” YouTube, 33:28, Doug Engelbart Institute, March 12, 2017. https://www.youtube.com/watch?v=hXdYbmQAWSM.

SRI International. “1968 "Mother of All Demos" with Doug Engelbart & Team (2/3).” YouTube, 32:42, Doug Engelbart Institute, March 12, 2017. https://www.youtube.com/watch?v=FCiBUawCawo.

SRI International. “Part 1 of 10: Engelbart and the Dawn of Interactive Computing: SRI's 1968 Demo (Highlights).” YouTube, 0:38, SRI International, December 11, 2008. https://www.youtube.com/watch?v=VScVgXM7lQQ&list=PLCGFadV4FqU2yAqCzKaxnKKXgnJBUrKTE&index=1.

SRI International. “Part 2 of 10: Engelbart and the Dawn of Interactive Computing: SRI's 1968 Demo (Highlights).” YouTube, 6:24, SRI International, December 11, 2008. https://www.youtube.com/watch?v=Xptc6f3Daoo&list=PLCGFadV4FqU2yAqCzKaxnKKXgnJBUrKTE&index=2.

SRI International. “Part 6 of 10: Engelbart and the Dawn of Interactive Computing: SRI's 1968 Demo (Highlights).” YouTube, 3:43, SRI International, December 11, 2008. https://www.youtube.com/watch?v=qisaqiysp2s&list=PLCGFadV4FqU2yAqCzKaxnKKXgnJBUrKTE&index=6.

Stanford University. “First Mouse.” accessed July 10, 2019. https://web.stanford.edu/dept/SUL/library/extra4/sloan/MouseSite/Archive/patent/Mouse.html.

Suhovecky, Mark, and Seth Anderson. “Beta Deployment: EaaSI in AWS or Bust!.” Software Preservation Network. Accessed July 6, 2019. https://www.softwarepreservationnetwork.org/blog/beta-deployment-EaaSI-in-aws-or-bust/.

Task Force on Archiving of Digital Information, Commission on Preservation Access, and Research Libraries Group. *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information*. Washington, D.C.: Commission on Preservation and Access, 1996.

Taylor, Hugh A. “The Media of Record: Archives in the Wake of McLuhan.” In *Imagining Archives: Essays and Reflections,* ed. Terry Cook, and Gordon Dodds, 63-74. Lanham:

Society of American Archivists and Association of Canadian Archivists in Association with Scarecrow Press, 2003.

The Consultive Committee for Space Data Systems. *Reference Model for an Open Archival System (OAIS).* Washington DC, June 2012.

University of Illinois Undergraduate Library. "About the Collection." accessed July 5, 2019. https://www.library.illinois.edu/ugl/gaming/about.

University of Michigan Library. "Computer & Video Game Archive." accessed July 5, 2019. https://www.lib.umich.edu/computer-video-game-archive.

University of Oxford Department of Computer Science. "Lung Chaste: developing software for multiscale lung ventilation simulations." accessed July 3, 2019. https://www.cs.ox.ac.uk/chaste/lung_index.html.

Upward, Frank, Sue McKemmish, and Barbra Reed. "Archivists and Changing Social and Information Spaces: A Continuum Approach to Recordkeeping and Archiving in Online Cultures." *Archivaria* 72 (2011): 197-237.

Upward, Frank. "In search of the continuum: Ian Maclean's 'Australian experience' essays on recordkeeping." in *The records continuum: Ian Maclean and the Australian Archives first fifty years, ed.* Sue McKemmish, and Michael Piggott, 110-130. Clayton: Ancora Press, 1994.

Upward, Frank. "Structuring the Records Continuum - Part One: Postcustodial Principles and Properties." *Archives & Manuscripts*, Vol. 24, Issue 2 (1996): 268-285.

Van Der Hoeven, J., R. Van Diessen, and K. Van Der Meer. "Development of a Universal Virtual Computer (UVC) for long-term preservation of digital objects." *Journal Of Information Science, 31*(3) (2005): 196-208.

Von Suchodoletz, Dirk, Klaus Rechert, and Isgandar Valizada. "*Towards Emulation-as-a-Service – Cloud Services for Versatile Digital Object Access.*" *The International Journal of Digital Curation* vol. 8 issue 1, (2013): 131-142.

Wadlow, Thomas A. "The Xeorx Alto Computer." *BYTE Magazine*, vol. 6 no.9 September 1981, 58, Internet Archive, accessed July 5, 2019, https://archive.org/details/byte-magazine-1981-09/page/n59.

Yeo, Geoffrey. "'Nothing is the same as something else': significant properties and notions of identity and originality." *Archival Science* 10, no. 2 (2010): 85-116.