Unruly Records:
Personal Archives, Sociotechnical Infrastructure, and Archival Practice

by

Chelsea Gunn

BA, Carlow University, 2008

MLIS, Simmons University, 2012

Submitted to the Graduate Faculty of the

School of Computing and Information in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

SCHOOL OF COMPUTING AND INFORMATION

This dissertation was presented

by

Chelsea Gunn

It was defended on

July 27, 2020

and approved by

Dr. Lara Putnam, Professor, Department of History

Dr. Tonia Sutherland, Assistant Professor, Department of Information and Computer Sciences, University of Hawai'i at Manoa

Dr. Stacy Wood, Assistant Professor, School of Computing and Information

Dissertation Director: Dr. Alison Langmead, Clinical Associate Professor and Director, Visual Media Workshop, History of Art and Architecture; Associate Professor, School of Computing and Information

Unruly Records:
Personal Archives, Sociotechnical Infrastructure, and Archival Practice

Chelsea Gunn

University of Pittsburgh, 2020

Personal records have long occupied a complicated space within archival theory and practice. The archival profession, as it is practiced in the United States today, developed with organizational records, such as those created by governments and businesses, in mind. Personal records were considered to fall beyond the bounds of archival work and were primarily cared for by libraries and other cultural heritage institutions. Since the mid-20th century, this divide has become less pronounced, and it has become common to find personal records within archival institutions. As a result of these conditions in the development of the profession, the archivists who work with personal records have had to reconcile the specific characteristics of personal materials with theoretical and practical approaches that were designed not only to accommodate organizational records but to explicitly exclude personal records.

These conditions have been further complicated by the continually changing technological landscape in which personal records are now created. As ownership of personal computers, access to the World Wide Web, and the use of networked social platforms have grown, personal records have increasingly come to be created, stored, and accessed within complex socio-technical systems. The infrastructures that support personal digital record creation today precipitate new methods and strategies, and an abundance of new questions, for the archivists who are responsible for collecting and preserving digital cultural heritage.

This dissertation considers how both the history of excluding personal records in the archival profession and the socio-technical systems that support contemporary personal record creation impact archival practice today. This research considers archival approaches to working with personal records created within three environments: personal computers, the open web, and networked social platforms. Ultimately, this dissertation seeks to reevaluate the role that personal records have previously occupied, and to center the personal in archival practice today.

Table of Contents

List of Tables

List of Figures

1.0 Introduction

This dissertation considers personal records as an archival genre, and investigates professional approaches to collecting and caring for born-digital personal records. Specifically, it seeks to explore the impact on archival practice of the socio-technical infrastructures in which personal records are created. In recent decades, there has been a growing recognition of the value of, and need to preserve, our digital cultural heritage, much of which is created and stored within remote, commercially controlled, technological systems. The proceeding research takes a historical approach, beginning with the exclusion of personal records in the historical development of the archival profession before examining the effects of three socio-technical information environments on practical and ethical approaches to collecting personal digital records.

Personal records have long occupied a complicated and, at times, contested space within archival theory and practice. The foundations of the archival profession as it is practiced in the United States today are grounded in the public archives tradition, which understands archives primarily as the generated within organizations, including government bodies and private businesses, in the course of conducting their affairs.[1] Personal records – those created by private individuals and families, unrelated to official business – were considered in this historical practice to be beyond the archival purview, and were left instead to the care of libraries, museums, and other cultural heritage institutions.[2] In spite of this, personal records are commonly housed within archival institution, including corporate, government, and university archives, particularly since

[1] Robert Fisher, "In Search of a Theory of Private Archives: The Foundational Writings of Jenkinson and Schellenberg Revisited," *Archivaria* 67 (2009): 5.
[2] Lester J. Cappon, "Historical Manuscripts as Archives: Some Definitions and Their Application," *The American Archivist* 19, no. 2 (1956): 101.

the mid-20th century.[3] The result, as Adrian Cunningham has suggested, has been that archivists responsible for collecting and caring for personal archives "have scratched their heads and pondered exactly how to apply the lofty principles of Sir Hilary Jenkinson to the personal papers of individuals."[4]

Though their status as archives has historically been contested, personal records have long been recognized as meaningful forms of information and evidence. Specifically, personal records have been valued for their contributions to collective and cultural memory.[5] For those who study history "from the bottom up," an approach that focuses on the lived experiences of individuals and groups, personal records can offer rich documentation of and reflection on events as they have been experienced and felt by the people directly affected by them.[6] The inclusion of the records of individuals within institutional archives bestows value and authority on those records, and can function as a method of elevating voices that have been historically under- or unrepresented in historical accounts.[7] Collecting and preserving personal records is one way archivists can preserve a record of our cultural heritage that does not only reflect the voices of individuals and institutions in positions of authority. In "Evidence of Me," a touchstone work for archivists concerned with personal records, Sue McKemmish articulated the transformative process through which the personal records of individuals can, together, contribute to a collective historical record: "a personal archive considered to be of value to society at large is incorporated into the collective

3 Luke J. Gilliland-Swetland, "The Provenance of a Profession: The Permanence of the Public Archives and Historical Manuscripts Traditions in American Archival History," *The American Archivist* 54, no. 2 (1991): 161.
4 Adrian Cunningham, "Beyond the Pale? The 'Flinty' Relationship Between Archivists Who Collect the Private Records of Individuals and the Rest of the Profession," *Archives and Manuscripts* 24, no. 1 (1996): 23.
5 Anthea Josias, "Toward an Understanding of Archives as a Feature of Collective Memory," *Archival Science* 11, no. 1–2 (2011): 95–112, https://doi.org/10.1007/s10502-011-9136-3.
6 Bonnie S. Brennan, *Qualitative Research Methods for Media Studies* (New York: Routledge, 2017), 95.
7 Terry Cook and Joan Schwartz, "Archives, Records, and Power: From (Postmodern) Theory to (Archival) Performance," *Archival Science* 2, no. 3–4 (2002): 171–85, https://doi.org/10.1007/BF02435620.

archives of the society, and thus constitutes an accessible part of that society's memory, its experiential knowledge and cultural identity;" preserved within the archives, "evidence of me" can become "evidence of us."[8]

Even as personal records have increasingly been accepted as valuable forms of evidence, both beyond and within the archival profession, debate about the definition, roles, and appropriate treatments of personal archives have persisted. They have been further complicated by technological changes that inform the production of personal records. Advancements in personal computing technologies, including the development of electronic mail, the World Wide Web, and social media platforms, have provided more people with the means not only to create but to share their own documents, whether with specific recipients or with the general public on the open web. Archivists have recognized the value of this seemingly endless trove of personal records – of so much more "evidence of us."[9] And as more personal records have been created in digital formats, with proprietary software, and within remote, commercially controlled platforms, many have expressed the sense that it is increasingly urgent to transfer these records to memory institutions, where they can be preserved and made accessible in the long term.

Concerns about the potential for format and platform obsolescence that threaten the longevity of these digital cultural heritage materials have compelled some archivists to adopt new proactive collecting practices, as well as technologies and tools designed to collect records created in new formats and environments. In some instances, as this dissertation will address, personal records have been collected directly from the commercial services in which they were created,

[8] Sue McKemmish, "Evidence of Me," *The Australian Library Journal* 45, no. 3 (January 1996): 175.
[9] Neil Beagrie, "Plenty of Room at the Bottom? Personal Digital Libraries and Collections," *D-Lib Magazine* 11, no. 6 (2005). http://www.dlib.org/dlib/june05/beagrie/06beagrie.html.

bypassing the involvement of individual record creators and further complicating the relationship between personal and organizational records.

Because the archival profession was developed with organizational, rather than personal, records in mind, archivists who work with personal records have had to consider how these unique sources of evidence and information can be served by archival theory and practice, and when new approaches must be developed. These theoretical and practical challenges are exacerbated by developments and changes in the infrastructures that support personal record creation.

1.1 Significance of Study

This dissertation begins by exploring the history of personal records within the archival profession in order to contextualize the current state of archival practice as it pertains to personal digital records, specifically. This research asks, first, how contemporary approaches to collecting personal digital records has been informed by the history of excluding personal records in archival theory and practice. Within this historical context, this research then asks how the socio-technical infrastructures that support personal record creation further impact archival approaches to working with personal records. Recognizing that personal digital records have been created in myriad complex systems, this dissertation considers three primary categories of personal digital records: those created and stored on local devices; those created and stored on the open web; and those created and stored within social network sites. Through this line of inquiry, the dissertation asks how these infrastructures might prompt a reexamination of the efficacy of traditional definitions of personal records, and how archivists can collect them in ways that enact an ethics of care.

As this research will demonstrate, personal archives have always been complicated. However, modern socio-technical infrastructure has rendered them complicated in new ways, which warrant further discussion. Just as archivists have worked to adapt professional policies and procedures that were designed for organizational records to serve personal records, so too have they had to adapt practices that were designed to suit paper records to serve those created in digital environments. This research takes place at the intersection of these two legacies of translation, adaptation, and reimagining. This dissertation is intended to contribute to ongoing discussions about ethics in personal archives and the role of the archivist in shaping cultural memory.

Since the mid-20th century, a steadily growing number of personal records have been created in digital formats. Files stored on personal computers, hard drives, and floppy disks presented new challenges to archivists: in particular, how to preserve and provide ongoing access to formats prone to rapid technological obsolescence.[10] By the late 20th century, the ability of a growing number of individuals to create personal records not only on personal computers but on the World Wide Web and, later, in social media platforms meant that personal records were not only born-digital, but created in remote, networked systems. In each of these environments, personal records are inextricable from the commercial programs and platforms in which they are created, stored, and accessed. For decades, archivists have worked to ensure that these digital records can be preserved for long-term access.

While digital preservation remains an ongoing concern, some have suggested that too much attention has been paid to this area, and that a preoccupation with the technical aspects of digital preservation has eclipsed other vital aspects of the archival endeavor, including appraisal and

[10] Jeff Rothenberg, "Ensuring the Longevity of Digital Information," *International Journal of Legal Information* 26, no. 1-3 (1998): 1-22.

accessioning practices.[11] Indeed, contemporary technological infrastructures have presented archivists with a host of new questions and challenges that merit further investigation. These include, but are not limited to: What does it mean to collect a personal record from a hard drive, from a remote server, or from a social media platform? How do we determine ownership of personal records that are created within commercial platforms? What is the relationship between ownership and control? What is the relationship between storage and control? How does the creation of records in commercial, networked spaces alter our understanding of privacy and publicity in personal archives?

When individuals create digital records, they create records that are precarious by nature. The causes of this precariousness are not strictly technological. These records – and, ultimately, the platforms in which they are created and stored – are highly dependent upon a host of technological, societal, and economic factors, each of which have the potential to impact the survival of the record long before archivists may even have the opportunity to acquire these personal records in the first place. My research interests have been sparked by moments in which the complexities of contemporary socio-technical systems have been made suddenly, unexpectedly more visible. These complexities often become more visible, as Steven Jackson has suggested, in moments of breakdown or destruction. Jackson calls this perspective "broken world thinking" – a way of approaching the study of infrastructure by asking "what happens when we take erosion, breakdown, and decay, rather than novelty, growth, and progress, as our starting points in thinking about the nature, use, and effects of information technology and new media."[12] Taking a similar

[11] Michael Moss and Tim Gollins, "Our Digital Legacy: An Archival Perspective," *Journal of Contemporary Archival Studies* 4, no. 2 (2017).
[12] Stephen J. Jackson, "Rethinking Repair," in *Media Technologies: Essays on Communication, Materiality, and Society* (Cambridge, MA: MIT Press, 2014): 221.

stance, media theorist Wolfgang Ernst has suggested that "only at a moment of technological breakdown will the medium become visible."[13] This dissertation draws upon these perspectives, examining the impact of precarious digital infrastructures on archival practice.

For decades now, archivists have collected the digital records of individuals, developing sophisticated strategies to preserve and provide access to these fragile materials. As the proceeding document will demonstrate, the preponderance of digital records, coupled with concerns about technological obsolescence, have prompted notable shifts in archival practice. In the following chapters, approaches to collecting personal records on local storage media, on the open web, and from social media platforms, will be examined.

1.2 Research Questions

This dissertation will address two primary research questions.

Question 1:

How have personal records come to be defined and understood throughout the history of the archival profession?

Question 2:

How do the specific socio-technical infrastructures supporting personal digital record creation impact archival approaches to collecting and preserving these materials?

[13] Wolfgang Ernst, *Digital Memory and the Archive* (Minneapolis: University of Minnesota Press, 2013): 49.

1.3 Themes of Study

This dissertation approaches the study of personal digital records by situating them within a series of ongoing developments in both the archival profession and the environments in which personal records are created. This research is prompted by questions about how the complex challenges posed by these socio-technical infrastructures require archivists to adapt and develop existing practices. This research does not focus on digital preservation, but instead on how concerns about technological obsolescence, failure, and destruction have galvanized action in other areas of archival practice. Using this lens, I approach my study of the impact of socio-technical infrastructure on approaches to working with personal papers by considering the relationship between three themes: Materiality; Custody and Control; and Privacy and Publicity. In the remainder of this section, I describe these themes and their function within this research.

1.3.1 Materiality

In his opening address to the Library of Congress's Digital Preservation 2014 conference, Matthew Kirschenbaum proposed a framework for moving beyond metaphorical descriptions of the digital in order to more accurately understand and examine software as, to borrow his phrase, "a thing;" or more specifically, to understand software as "a logical, spatial, and imaginative artifact, subject to craft and technique, to error and human foible."[14] Descriptions of digital artifacts too often rely upon rhetorical turns of phrase that ultimately serve to remove digital artifacts from the physical, material world. Whether one is describing "cyberspace" or the storage of data "in the

[14] Matthew Kirschenbaum, "Software, It's a Thing," *Matthew Kirschenbaum (Medium)*, July 25, 2014, https://medium.com/@mkirschenbaum/software-its-a-thing-a550448d0ed3.

cloud," these turns of phrase relocate our digital artifacts somewhere beyond bodily, physical artifacts and experiences. This dematerializing effect is not limited to online artifacts and spaces; indeed, writing about word processing documents, digital photographs, and software stored locally on a local personal computer has allowed for a rhetorical distancing between what one views on a screen and the digital artifacts that render it visible to the human eye. In this context, Ciaran B. Trace has likewise called for archivists to "open the computer's black box and become as familiar with computer hardware, application software, system software, and firmware as we have become with earlier writing, recording, and storage technologies."[15] Trace makes a compelling argument that by approaching digital technologies with a student's curiosity, archivists stand a much greater chance of understanding digital records as both conceptual objects ("records as they are presented and viewed by the user") and physical objects ("things that are inscribed on a medium").[16] By reorienting attention to the material and spatial characteristics digital objects, archivists may achieve a more precise understanding of digital records and the infrastructures that support them, both prior to and after their entry into the archives.

Trace, Kirschenbaum, and others have advised their audiences to avoid the trappings of screen essentialism, the tendency to define a digital object by what is seen on the screen, and ultimately obfuscating the material and logical components that remain unseen.[17] These points are

[15] Ciaran B. Trace, "Beyond the Magic to the Mechanism: Computers, Materiality, and What It Means for Records to Be 'Born Digital,'" *Archivaria* 72 (2011): 5.
[16] Trace, "Beyond the Magic to the Mechanism," 8.
[17] Nick Montfort and Matthew Kirschenbaum have written about the concept of "screen essentialism," which refers to a "prevailing bias" toward display technologies, eliding the functions and, perhaps most importantly, the materiality of digital objects. Trevor Owens has explored screen essentialism from a digital preservation perspective, making the case for the utility of this concept for "understanding the integrity of digital objects." Montfort, "Continuous Paper: The Early Materiality and Workings of Electronic Literature," *MLA Convention*, December 28, 2004, https://nickm.com/writing/essays/continuous_paper_mla.html; Kirschenbaum, *Mechanisms: New Media and the Forensic Imagination* (Cambridge, MA: MIT Press, 2008), 31; Owens, "Glitching Files for Understanding: Avoiding Screen Essentialism in Three Easy Steps," *The Signal*, November 5, 2012, https://blogs.loc.gov/thesignal/2012/11/glitching-files-for-understanding-avoiding-screen-essentialism-in-three-easy-steps/.

well taken, and yet, in the context of remotely located, networked information culture, "opening up the black box" becomes exponentially more complicated. In addition to requiring some understanding of the personal computer, hard drive, or other discrete object, taking a look behind the screen to better understand our digital records now requires some knowledge of the often invisible or hard-to-see sociotechnical infrastructure which supports networked record creation and remote storage.[18] Further still, it often requires access to these infrastructures. This access is increasingly hard to obtain as our records become further entangled in remote, corporate ecosystems, a condition that has clear implications for archivists seeking to acquire custody of those records.

The collections discussed throughout this dissertation will be described with particular attention to their material – meaning, as Kirschenbaum suggests, their physical, but logical, spatial, and imaginative – characteristics. This conceptualization of materiality will rely upon evidence of the physical contours and geographic location of digital records, as well as on documentation of how these artifacts are crafted and cared for. The material characteristics of digital records, this dissertation suggests, play a central role in the archival approaches to working with these records, and in particular, to obtaining custody and control of them.

1.3.2 Custody and Control

Within the archival field, the term custody refers to the "care and control" of records, especially for the purposes of security and preservation.[19] In archival theory and practice, the

[18] Maureen Henninger and Paul Scifleet, "How Are the New Documents of Social Networks Shaping Our Cultural Memory," *Journal of Documentation* 72, no. 2 (2016): 277–98, https://doi.org/10.1108/JD-06-2015-0069.
[19] "Custody," *Dictionary of Archives Terminology*, https://dictionary.archivists.org/entry/custody.html.

custodial history of a collection, meaning "the succession of offices families, or persons who held materials from the moment they were created," can provide essential information about the integrity, authority, and context of the collection's contents.[20] Custody is a broad term, possessing both physical and legal meanings. Importantly, maintaining physical custody of records does not automatically signify legal ownership of them. It is therefore of great importance that the conditions governing both physical and legal custody be established in a gift agreement between donor and repository.[21] These terms, often already complex, can be even more so in the context of born-digital records, stored remotely in commercially-controlled platforms.

Many individual record creators have found it difficult to understand the terms that dictate the control of their own personal digital records. In their 2017 study of user attitudes about intellectual property and personal digital archives, Catherine C. Marshall and Frank M. Shipman found that expectations and social norms around ownership were wildly divergent among their participants, and that within contemporary networked contexts such as the World Wide Web, it was challenging for record creators to confidently assert their own control or ownership, in large part due to the awareness that any of their records "may be associated with a virtual web of stakeholders."[22] This dissertation takes the work of identifying stakeholders and understanding their various claims to personal digital records as a central concern.

The specific systems in which personal records are created have a direct impact on the ways in which custody of those records can be established. In many instances, the methods used

[20] "Custodial History," *Dictionary of Archives Terminology*, https://dictionary.archivists.org/entry/custodial-history.html.
[21] "A Guide to Deeds of Gift," Society of American Archivists, https://www2.archivists.org/publications/brochures/deeds-of-gift#.V0kj984sj6U.
[22] Catherine Marshall and Frank Shipman, "Who Owns the Social Web?" *Communications of the ACM* Vol. 60. New York: ACM, (2017): 52.

to transfer digital records into archival custody – including migration, web archiving, and social media API archiving – are transformative, and have been developed in response to the degree to which personal digital records are accessible to archivists. Within this state of affairs, archivists have had to ask a number of questions about the control of personal digital records. These questions include, but are not limited to: When one creates a personal record in a commercial platform, how are ownership and custody determined? How does one establish control or assert ownership over a record that is stored on a service provider's server, in a remote, unknown location? Particularly when collecting records that have been created in remote, commercial platforms, archivists have been required to work within the bounds of the platforms' Terms of Service. As this research will demonstrate, those terms of service are designed to protect the interests of the platform, rather than its users. This will have a direct impact on ethical dimensions of archival work, including, notably, issues related to privacy in personal archives.

1.3.3 Publicity and Privacy

The concept of privacy is complicated by socio-technical systems, and in particular by networked social platforms, within which many personal records are considered by default to be public. However, discussion and debate on the topic of privacy have always been integral to the study, acquisition, and preservation of personal records. In *The Ethical Archivist*, Elena Danielson has suggested that "the violation of privacy is an intrinsic and unavoidable part of archival work, because it involves the secondary use of documents, which were created for another, so called primary, purpose."[23] Indeed, the Society of American Archivists defines privacy as "the quality or

[23] Elena S. Danielson, *The Ethical Archivist* (Chicago: Society of American Archivists, 2010).

state of having one's personal information or activities protected from unauthorized disclosure by another."[24] The distinction between authorized and unauthorized users is one that is essential to the proceeding research, which will resist the notion that "public" and "private" are binary, oppositional concepts. Instead, this dissertation considers the ways in which privacy is dependent upon context, and the ways in which archival approaches to collecting and providing access to personal records constitute a shift from one context to another. Through this historical examination, this dissertation will examine how contemporary socio-technical infrastructure and archival collecting practices can at times upend traditional policies and procedures. In this landscape, archivists must ask when they themselves may be unauthorized users of personal records.

Helen Nissenbaum's scholarship on contextual integrity provides a conceptual framework for this dissertation's focus on privacy in personal archives. As Nissenbaum explains, individuals create and use personal records in specific contexts and feel that their privacy has been violated when their personal information is passed inappropriately from one context to another.[25] As several of the collections discussed throughout this research will attest, the very processes through which personal records are transferred from their sites of creation and use into archival spaces, have the potential to constitute such a violation. Contextual integrity offers a critical framework for approaching this research without resorting to simplistic explanations. Instead, this framework can be used to encourage archivists to learn about the specific contexts in which records are created, stored, and used, in order to better understand how archival processes may constitute a shift in context for those records.[26] This dissertation's exploration of personal digital records, whether

[24] "Privacy," *Dictionary of Archives Terminology*, https://dictionary.archivists.org/entry/privacy.html.
[25] Helen Nissenbaum, "Privacy as Contextual Integrity," *Washington Law Review* 79, no. 1 (2004).
[26] Nissenbaum, "Privacy as Contextual Integrity."

created and stored on local devices or online, in networked systems, relies upon a contextual understanding of privacy. In the proceeding research, personal records are understood to exist on a continuum of publicity and privacy and, it is important to note, their positions along this continuum are not fixed in a single point.

1.4 Conclusion

Using these concepts, this dissertation seeks to establish personal records as an archival genre. If personal records are, as has been repeatedly suggested throughout the history of western archival studies, fundamentally unique forms of evidence, and unlike organizational records, then what makes them so? Understanding personal records as an archival genre then requires a close examination of the contexts of their creation and use, and specifically, the sociotechnical infrastructures which support them.

This dissertation begins by considering the impact of the archival profession's exclusion of personal records. Working with theoretical and practical approaches designed to serve organizational records, archivists who collect and care for personal records have been required to adapt existing strategies and, in some instances, develop new ones altogether in their work. Though the value of personal records as historical evidence and records of collective memory is widely recognized, others have observed that the historical "sidelining" of personal records continues to have a meaningful impact on archival practice.[27] Within the context of this history of exclusion,

[27] Jennifer Douglas, "Getting Personal: Personal Archives in Archival Programs and Curricula," *Education for Information*, 33, no. 2 (2017): 89-105; Catherine Hobbs, "The Character of Personal Archives: Reflections on the Value of Records of Individuals," *Archivaria* 52 (February 2001): 126-135.

which will be discussed in greater detail in Chapter Three, this dissertation then considers the approaches taken to collecting personal records within a steadily changing technological landscape.

Using the thematic lenses of Materiality, Custody and Control, and Privacy and Publicity, this dissertation explores a range of approaches to collecting and caring for persona records created in various digital environments. Specifically, it will consider approaches taken to personal records created, and subsequently acquired by archivists, on local devices, on the open web, and within networked social platforms. This research suggests that the specific socio-technical infrastructure in which personal records are created has a direct and meaningful impact on the ways in which custody is transferred to or taken by archivists, and in turn, the ways in which personal records are understood to be, and treated as, private or public. A key objective of this work is the identification of emerging areas of concern for archivists responsible for collecting and preserving personal digital records within their institutions. The next chapter, Research Design, will detail the methodological approach, data sources, and structure of this dissertation.

2.0 Research Design

In the proceeding dissertation, I address my research questions by conducting a qualitative study. Qualitative research is well suited to questions on archives and archival practice, as it takes into consideration not only gathered data, but the original contexts from which the data are drawn.[28] Context is, of course, a central concern in archival scholarship and practice. This research is inductive and interpretivist. In the following pages, I do not seek to test a specific hypothesis, but instead to explore specific phenomena in order to contribute to a greater understanding of areas of interest and concern for archivists who work with personal digital records.

2.1 Genre Archaeology

I approach my research questions by undertaking what might be effectively described as a genre archaeology. Specifically, I approach the study of personal records, and personal digital records in particular, as an archival genre, within which one may expect to find many diverse formats, varieties of content, and modes of creation. Among the goals of this dissertation is the further elucidation of personal records within the archival field. An understanding of personal records as an archival genre builds upon the investigation of the material, historical, and sociocultural conditions that produce them. This work begins with a review of the archival literature on personal records, beginning with early theoretical and instructional works and

[28] G. E. Gorman and Peter Robert Clayton, *Qualitative Research for the Information Professional: A Practical Handbook* (Library Association Publishing, 1997): 3.

continuing through contemporary scholarship on personal digital archives. This review will serve to contextualize contemporary practice within the history of the archival profession, laying the groundwork for the study of professional archival approaches to collecting locally-stored, web-based, and socially networked personal digital records. These modes of analysis, applied together, will help to establish and describe personal records as an archival genre.

In their introduction to a special issue of *Archival Science* dedicated to the application of genre studies to archival work, Gillian Oliver and Wendy M. Duff observed that genre has not "figured prominently in either archival discourse or practice to date," though, in their estimation, the concept of genre has "the potential to yield much of relevance to the archival community."[29] Because genre has not factored significantly into archival studies, Oliver and Duff have suggested, it may often be misunderstood as referring to either specific types or formats of documents, or to the organizational classification structures used to assign literary works to broad, descriptive groups such as romance, mystery, or fantasy.[30] However, some archival scholars have attempted to establish a more productive conception of genre in archival studies. Indeed, the "recognition of context, and exploration of influences that shape and fashion communicative activity," that is found in contemporary genre studies, "resonates with the archival endeavor," as Oliver and Duff have argued.

In their own exploration of the utility of genre in the study of digital recordkeeping, Oliver, Yunhyong Kim, and Seamus Ross reviewed the genre studies literature to argue for documentary genre as an archival concept.[31] Genre, they have noted, tends more toward the fluid than the fixed,

[29] Gillian Oliver and Wendy M. Duff, "Genre Studies and Archives: Introduction to the Special Issue," *Archival Science* 12, no. 4 (2012): 373.
[30] Oliver and Duff, "Genre Studies and Archives," 373.
[31] Oliver, Gillian, Yunhyong Kim, and Seamus Ross. "Documentary Genre and Digital Recordkeeping: Red Herring or a Way Forward?" *Archival Science* 8, no. 4 (2008): 295-305.

and thus can be resistant to taxonomy and classification. However, they did find connections between genre and digital recordkeeping in the example of a taxonomy developed by Yoshioka et al., which they found to be "flexible enough to accommodate change and flux," while still providing structure.[32] The taxonomy was based on six dimensions that can comprise a genre – purpose, content, timing, location, participants, and structure and media – which map rather neatly to a structure that will be familiar to many – why, what, when, where, who, and how, respectively.[33] Gillian et al. noted that this taxonomy draws attention away from a singular understanding of genre that is focused on content or record type, in part because of its allowance for contextual data.[34] This dissertation research draws structural inspiration from the concept of genre as a flexible system for identifying themes in context.

Within the archival literature, researchers have offered other ways of meaningfully incorporating scholarship from genre studies. In their assessment of personal recordkeeping as a documentary genre, Pamela McKenzie and Elisabeth Davies have proposed that an approach to the study of personal records that pays particular attention to the circumstances of record creation aligns with "the North American approach to rhetorical genre analysis, which seeks to understand genres and groups of genres as constituents of social action."[35] Oliver and Duff have meanwhile suggested that we might think of genre as "a pattern of communication that conforms to community norms."[36] The proceeding research suggests that these patterns of communication and community norms are not only present in the production of personal digital records, but in the

[32] Oliver et al., "Documentary Genre and Digital Recordkeeping," 298.
[33] Gillian et al., "Documentary Genre and Digital Recordkeeping," 298.
[34] Gillian et al., "Documentary Genre and Digital Recordkeeping," 298.
[35] Pamela J. McKenzie and Elisabeth Davies, "Genre Systems and 'keeping Track' in Everyday Life," *Archival Science* 12, no. 4 (2012): 438.
[36] Oliver and Duff, "Genre Studies and Archives," 373.

practices employed by the archivists who collect them – the community norms of the archival profession.

2.2 Methodology

Recognizing that, as discussed in the preceding pages, personal digital archives are shaped significantly by their socio-technical contexts, I have attempted to conduct this research such that it allows for exploration of both technological and archival infrastructures. I have undertaken this study by employing mixed qualitative research methods. Specifically, I have conducted this study through the application of historical and thematic analysis and ethnography of infrastructure. Through the analysis of primary and secondary sources, I have observed and describe first the personal digital records themselves, and then the contexts in which personal digital records have been accessed, collected, and made accessible within memory institutions. Through the use infrastructure ethnography, I have studied personal digital records as cultural and material artifacts, from the infrastructures supporting their creation and use, to those in which they have been collected for long-term access. This dissertation approaches its subject – personal digital records – as a genre that has been continually shaped by its cultural and physical contexts. As Bonnie S. Brennan has written in *Qualitative Research Methods for Media Studies*, studying such a subject, or genre, requires that the researcher understands that:

> "all documents of material culture, including newspapers, books, films, popular music, television programs, comic strips, current fashions as well as newer media such as Facebook, Second Life and Twitter, are produced under specific

> conditions, and that any or all of these cultural products can provide us with insights about our society at a particular historical place and time."[37]

Records created in any of these mediums or systems have much to tell us about not only the individuals who created them, but about the social, political, technological, and economic contexts in which they were created. Personal digital records are shaped not only be cultural and social norms, but by the affordances and limitations of the technologies with which they have been created. The study of personal records as an archival genre draws upon inductive, descriptive research methods, with particular attention to context – in other words, Yoshioka, et al.'s taxonomical interpretation of why, what, when, where, who, and how.

In *The Practice of Social Research,* Earl Babbie outlined three common purposes of social research: exploration, description, and explanation.[38] Exploration, Babbie explained, is typically employed when the subject of study is still somewhat fresh or new.[39] This type of study can help a researcher to better understand a particular phenomenon, in addition to potentially satisfying a sense of personal curiosity. Arguably, most inductive research begins in precisely this way. Exploratory research allows the researcher to identify key details, variables, and paths forward. While a study, such as this one, may have its origins in a sense of personal curiosity, its greater objective is to describe the phenomena of study and identify potential paths forward for archivists who work with these records. To that end, producing a structured, well-documented, and reliable description of these records and how they are impacted by their various infrastructures is essential to this work. Though Babbie summarizes this rather simply ("the researcher observes and then describes what was observed"), to produce a trustworthy account of past events is far from a

[37] Brennan, *Qualitative Research Methods for Media Studies*, 2.
[38] Babbie, *The Practice of Social Research, 14th edition* (Boston, MA: Cengage Learning), 90.
[39] Babbie, *The Practice of Social Research,* 90.

simplistic endeavor.[40] The methods outlined in the following sections provide structured modes of exploration and description, through which phenomena can be examined, and patterns and implications might be observed and interpreted.

2.2.1 Historical and Thematic Analysis

Philip Gardner has described historical analysis as the critical analysis of primary sources in order to make sense of the past.[41] Historical analysis is often used in combination with other research methods, as it is within in this dissertation, in order to address research questions in the social sciences. Indeed, Gardner suggested, historical analysis might be well understood as a "pervasive and necessary technique in its own right, without which no account of phenomena in the present may be properly understood."[42] The proceeding research takes this assessment to heart, operating under the perspective that contemporary treatments of personal digital archives are best understood within a historical narrative of personal archives throughout the history of the modern archival profession.

Within library and information science specifically, Charles Busha and Stephen Harter have suggested that a scientific approach may be used in historical inquiries, which they described as "the systematic recounting of past events pertaining to the establishment, maintenance, and utilization of systematically arranged collections of recorded information or knowledge."[43] They do, however, note that not all researchers support the proposition of historical research as a

[40] Babbie, *The Practice of Social Research,* 91.
[41] Philip Gardner, "Historical Analysis," in *SAGE Dictionary of Social Research Methods* (London: SAGE Publications, 2006).
[42] Gardner, "Historical Analysis," 135.
[43] Charles Busha and Stephen P. Harter, *Research Methods in Librarianship: Techniques and Interpretations,* (New York, NY: Academic Press, 1980): 93.

scientific method, as it "lacks the means of achieving rigorous analytical precision and precise explanations."[44] In response, Robert F. Berkhofer, Jr. has suggested that this does not mean that historical research within the social sciences cannot provide "rigorous analytical precision," only that is often difficult to achieve that standard.[45] Taking this into consideration, this dissertation seeks to use historical methods to "work toward *understanding* [the archival profession] as it existed during a specified period of time," using systematic historical methods to identify, document, and analyze the practices and discourses used to preserve personal digital records. The methods and interpretive lens of a cultural historian, specifically, are useful in this endeavor.

Brennan has summarized the approaches of two types of historical researchers: traditional historians and cultural historians. As she explains, a traditional approach to history presents "narratives as fact-based objective explanations of events, issues, and problems," whereas a cultural approach to historical research accepts that "researchers cannot remain neutral about historical evidence – that they interpret the past using relevant concepts and theories in order to understand the evidence that they are able to access."[46] This is an important distinction that must be acknowledged in advance of this research. This dissertation identifies, collects, and analyzes historical artifacts and documentation to produce one interpretation; a different researcher working in a different time, or from a different discipline, might produce another interpretation altogether. This does not, however, constitute a methodological weakness, but instead reflects the influence of social constructivism as an interpretive framework.

To this point, Brennan has explained that cultural media history "emphasizes the collective process of people connected with communication within specific economic, political, and cultural

[44] Busha and Harter, *Research Methods in Librarianship*, 91.
[45] Robert F. Berkhofer, Jr., *A Behavioral Approach to Historical Analysis.* New York: Free Press, 1969: 7.
[46] Brennan, *Qualitative Research Methods for Media Studies*, 95.

environments."[47] Busha and Harter have likewise suggested that historical research in the information sciences considers the "social, economic, political, intellectual, and cultural environments" in which events have occurred in order to relate the causes and effects of those events.[48] I have applied this methodological approach in order to study both the personal digital records themselves and the efforts made by archivists to preserve them. This is accomplished by examining each of these subjects within their contemporaneous social, economic, and technical contexts. This focus is vital to the goals of this dissertation, in part because, as Brennan suggests, a focus on only the technological challenges of preserving personal digital records would "privilege the tools, making the technologies seem more important than those who use them."[49]

While the myriad technological threats to digital artifacts are certainly worthy of scholarly attention, researchers and practitioners in the archival field have observed that significant attention and resources already been dedicated to these aspects of the contemporary archival landscape. In a recent article, Michael S. Moss and Tim J. Gollins have suggested that such an overemphasis has been placed on digital preservation in recent years that it has ultimately been to the detriment of holistic archival practice with regard to modern digital records, and in particular, to the decline in scholarship dedicated to appraisal.[50] In the proceeding dissertation, it is my objective to address this gap in the scholarship, focusing on a broader depiction of archival practices with regard to personal digital records, and centering people within this work.

My approach to this research draws additionally on thematic analysis, a form of narrative analysis. Catherine Kohler Riessman has written about narrative analysis, explaining that within

[47] Brennan, *Qualitative Research Methods for Media Studies*, 96.
[48] Busha and Harter, "Research Methods in Librarianship," 93.
[49] Brennan, *Qualitative Research Methods for Media Studies*, 96.
[50] Moss and Gollins, "Our Digital Legacy: An Archival Perspective."

this approach to research, investigators, "select and organize documents, compose field notes, and/or choose sections of interview transcripts for close inspection."[51] A thematic approach to this work is "useful for theorizing across a number of cases – finding common thematic elements" throughout them.[52] This approach is well-suited to this dissertation, which examines a variety of cases in each chapter in an effort to identify overarching themes across all collections studied.

As discussed further in the Data Sources section that comes in the following pages, much of this research relies on the examination of personal records as digital artifacts themselves, in addition to investigate my research questions by examining primary and secondary documents related to their social, economic, and political, contexts. In order to do this, I complement my historical analysis with methods from infrastructure studies.

2.2.2 Ethnography of Infrastructure

Defined broadly, ethnography is a methodological approach that is dedicated to "understanding what people believe and think, and how they live their daily lives."[53] Long associated with the field of anthropology, it has been taken up in other disciplines as well. Ethnography often involves long-term, immersive field research, in which the researcher observes activities, behaviors, and environments of specific groups of people. As Sylvain Cibangu has written, ethnographic research "requires the researchers to immerse themselves in the real world of the selected phenomenon or topic in order to (re-)enact the fullness of lifeworlds into the senses

[51] Catherine Kohler Riessman, "Narrative Analysis," in *SAGE Dictionary of Social Research Methods* (London: SAGE Publications, 2006): 186.
[52] Riessman, "Narrative Analysis," 187.
[53] Brennan, *Qualitative Research Methods for Media Studies*, 159.

of the reader."[54] Though ethnography has been traditionally employed in the study and observation of human subjects, it has also recently been applied effectively to the study of media, information systems, and infrastructure themselves, in addition to the ways in which they are used or interacted with by a group of people.

Information infrastructures can be challenging to study or analyze because they are often, almost by definition, invisible. In *Sorting Things Out*, Geoffrey Bowker and Susan Leigh Star observed that "the easier [effective information systems] are to use, the harder they are to see. As well, most of the time, the bigger they are, the harder they are to see."[55] In "The Ethnography of Infrastructure," Star further acknowledged that many aspects of infrastructure may initially seem to be uninteresting or mundane, so often manifested as "lists of numbers and technical specifications, or as hidden mechanisms subtending those processes more familiar to social scientists."[56] However, these hidden, seemingly mundane systems have the potential to impact many aspects of our daily lives. This dissertation seeks to illuminate some aspects of the infrastructures supporting personal digital record creation and collection, using broken world thinking as a lens for viewing those infrastructures. As Bowker and Star suggest, infrastructure is at its most invisible when it is operating as intended. However, using Jackson's approach, it is possible to glimpse infrastructure at moments of breakdown. In those moments, there is an opportunity to, as Star has suggested, "unearth the dramas inherent in system design creating, to restore narrative to what appears to be dead lists."[57] Through systematic examination of the

[54] Sylvain K. Cibangu, "A Memo of Qualitative Research for Information Science: Toward Theory Construction," *Journal of Documentation* 69, no. 2 (March 1, 2013): 202.
[55] Geoffrey C. Bowker and Susan Leigh Star, *Sorting Things Out: Classification and Its Consequences*. (Cambridge: MIT Press, 1999): 33.
[56] Susan Leigh Star, "The Ethnography of Infrastructure," *American Behavioral Scientist* 43, no. 3 (1999): 377.
[57] Star, "The Ethnography of Infrastructure," 377.

"architecture and use" of information systems, we are able to better understand how "individuals and communities meet infrastructure."[58] This makes ethnography of infrastructure an apt method for this dissertation's goals of interrogating sociotechnical systems.

Bowker and Star have acknowledged that "infrastructures are never transparent for everyone, and their workability as they scale up becomes increasingly complex."[59] Particularly because this dissertation concerns proprietary services and archival institutions, each with complex internal workings and documentation, systems, and norms that remain private for myriad reasons, there will be aspects of the infrastructures examined that will remain invisible. However, within this research project, any gaps, limits, and pressure points within these systems that are identified in the course of research can also function as findings.

2.3 Data Sources

As discussed in the preceding section, this dissertation's methodology includes historical, narrative, and ethnographic methods, and its units of analysis are social and cultural artifacts – specifically, personal digital records that have been collected by memory institutions. Historical analysis relies upon data sources which are found or encountered, rather than produced, by the researcher.[60] The earliest stages of the research process thus included an ongoing environmental scan of personal digital records within memory institutions with the purpose of identifying collections for closer examination.

[58] Bowker and Star, *Sorting Things Out*, 33.
[59] Bowker and Star, *Sorting Things Out*, 33.
[60] Gardner, "Historical Analysis," 135.

Drawing boundaries in historical research can be a daunting undertaking. The list of potential data sources has the potential to expand continually in the course of research as new resources and information are gradually revealed. Gardner notes that the "traces" studied in historical analysis can, after all, take many shapes; "ranging from everyday ephemera, artefacts and visual images, to old buildings, archaeological sites or entire landscapes."[61] Most often used, in Gardner's assessment, however, are "written documents, whether of public or private origin."[62] Sources used throughout this dissertation do, in fact, consist primarily of written documents, including digital artifacts such as websites and social media platforms, technical documentation, news and popular media coverage, and existing research and archival documentation. Secondary sources, such as research by scholars from archival and information science and other relevant fields and coverage of collections in newspapers, magazines, and blogs will also be drawn on for context.

While the focus of my research is on born-digital personal records, the historical framing of my dissertation necessitates that I situate those records within historical archival traditions of working with both non-digital personal records, as this history has informed the policies and procedures still in use today. This research will thus draw on perspectives related to print and other non-digital personal records valuable points of comparison and contextualization.

[61] Gardner, "Historical Analysis," 135.
[62] Gardner, "Historical Analysis," 135.

2.3.1 Criteria for Inclusion

This dissertation's objects of study constitute a collection of purposeful, comparative case studies. As Jupp explains, this method entails "the selection and analysis of cases that are similar in known ways and differ in other ways, with a view to formulating or testing hypotheses."[63] The cases examined in the following pages are those which include digital records created by private individuals, and which have been subsequently acquired or collected for long-term preservation and access within memory institutions. They have been selected partially, as Jupp describes, on the basis of their known similarities: the cases in each chapter include personal digital records created in the specific socio-technical infrastructure being examined. Each chapter will consider how these collections are similar, as well as how they differ from one another.

The sites of study selected for this dissertation serve as examples of modern archival approaches to collecting and working with personal digital records. This research operates from a perspective that because personal records by nature idiosyncratic and characterized by a "wildness," to borrow Catherine Hobbs' phrase, wholesale generalization is not a goal.[64] Rather, the intention of this research is to highlight cases which point to ways in which archival practice, as it pertains to personal digital records, might be productively reexamined.

Collections studied in this dissertation share qualities that have been explored in the scholarship of the Documentation movement, and in particular the work of Suzanne Briet and

[63] Victor Jupp, "Comparative Method," in *SAGE Dictionary of Social Research Methods* (London: SAGE Publications, 2006): 33.
[64] Catherine Hobbs, "Reenvisioning the Personal: Reframing Traces of Individual Life," in *Currents of Archival Thinking* 2nd ed. (Santa Barbara: Libraries Unlimited, 2010), 213.

Michael Buckland. As Buckland summarizes them, Briet's criteria for determining document status are well suited to this purpose. They are as follows:

> "(1) There is materiality: Physical objects and physical signs only;
> (2) There is intentionality: It is intended that the object be treated as evidence;
> (3) The objects have to be processed: They have to be made into documents; and, we think;
> (4) There is a phenomenological position: The object is perceived to be a document."[65]

The records examined in this dissertation are all material; indeed, their materiality is one of the primary themes of the study. They are also those which have been accessioned by archives or other memory institutions. They are preserved for their evidential value, processed by professionals into archival objects considered to be worthy of study. They are also those for which the documentation of archival and preservation actions taken have been made accessible to the public. The contents of these collections exemplify difference as well as similarity. They have been selected in part because they demonstrate a range of motivations and strategies for acquisition. In some instances, these personal records have come to the archives after years of benign neglect, stored on personal computers; in some, they have been proactively collected by archivists as a result of impending destruction or loss.

As Jupp notes, historical analysis may itself often employ comparative methods, such as in research that compares one time period to another.[66] While the research conducted in this dissertation is not engaged in the direct comparison of specific time periods, it does compare

[65] Michael K. Buckland, "What is a "Document?" *Journal of the American Society for Information Science* 48, no. 9 (1997): 806.
[66] Jupp, "Comparative Method," 33.

collections that include personal records that have been created within three distinct socio-technical environments over time. The cases explored in this dissertation have been chosen in part for their comparative potential. They possess characteristics common to personal digital records created in each socio-technical environment. However, as Chapter Three will discuss, the individualized nature of personal records makes generalization a challenging, perhaps even flawed, objective.

2.3.2 List of Collections

Each of this dissertation's primary content chapters studies five collections, identifying similarities, differences, and overarching themes in their Materiality, Custody and Control, and Privacy and Publicity. This section provides a list of the collections to be discussed throughout each of those chapters. A descriptive overview of each of these collections can be found in Appendix A.

Table 1 List of collections discussed in Chapter Four

Collection Title	Collecting Institution
Deena Larsen Collection	Maryland Institute for Technology in the Humanities, University of Maryland
Rafael Fajardo Digital Materials	Media Archaeology Lab, University of Colorado Boulder
Salman Rushdie Papers	Manuscripts, Archives, and Rare Books Library, Emory University
Susan Sontag Papers	Library Special Collections, University of California, Los Angeles
Toni Morrison Papers	Manuscripts Division, Department of Special Collections, Princeton University

Table 2 List of collections discussed in Chapter Five

Collection Title	Collecting Institution
GeoCities Collection	Archive Team
GeoCities Special Collection 2009	Internet Archive
Katie Lee Collection	Special Collections and Archives Department, Northern Arizona University
Mormon Missionary Collection	L. Tom Perry Special Collections, Brigham Young University
Zine Web Archive	Library of Congress

Table 3 List of collections discussed in Chapter Six

Collection Title	Collecting Institution
Confederate Monument Protest Collected Tweets	University Archives, University of North Carolina at Chapel Hill
#MeToo Collection	Schlesinger Library, Harvard University
Thomas S. Mullaney Papers	University Archives, Stanford University
Twitter Archive	Library of Congress
Ursula K. Le Guin Papers	Special Collections and University Archives, University of Oregon

2.4 The Researcher's Perspective

In *The Sage Dictionary of Social Research Methods*, Maggie Sumner described qualitative research as being grounded in interpretivism and inductivism.[67] This dissertation takes an inductive approach, beginning with specific objects of study in an attempt to identify larger trends or issues. This research resists positivist assumptions, and instead considers the ways in which data sources

[67] Maggie Sumner, "Qualitative Research," *The SAGE Dictionary of Research Methods* ed. Victor Jupp (London: SAGE Publications, 2006), 249.

and examples studied are socially constructed. It is with these conceptual frameworks that I approach this dissertation, always mindful of the ways in which my own experiences as both a professional archivist and a creator of my personal own personal archive have the potential to inform my research.

By applying the methods of a cultural historian, I align myself with the postmodern movement in archival scholarship, which rejects the notion that archival practice or archives themselves can provide neutral or objective representations of the past.[68] Archival practice and documentation, much like record creation, are informed by the individual and by the context in which they are working, and are therefore not neutral. This approach is supported by the use of ethnographic and narrative methods. In their examination of narrative inquiry in archival work, Barbara Morgan-Fleming et al., ask several questions that underpin my approach historical analysis: "Whose story is it? Who authored this tale? Whose voices were included? Whose voices were silenced?"[69] These questions motivate the research in the proceeding document.

2.5 Structure of the Dissertation

The Introduction and Research Design in the previous pages have established my research questions and my approach to addressing them. This dissertation is the result of years working with personal records, both within and outside of archival institutions. These experiences have inspired a number of questions that will be explored throughout the chapters that follow, such as:

[68] Terry Cook, "Archival Science and Postmodernism: New Formulations for Old Concepts," *Archival Science* 1 (2001): 3-24.
[69] Barbara Morgan Fleming, Sandra Riegle, and Wesley Fryer, "Narrative Inquiry in Archival Work," in *Narrative Inquiry: Experience and Story in Qualitative Research*. San Francisco: Jossey Bass Publishers (2000): 81-98.

What is a personal record? How have personal records been defined and understood by archivists over time? Is the archival treatment of personal records altered by their location or site of storage, and if so, how? How do the challenges of collecting and preserving born-digital networked personal records necessitate or inspire a reexamination of historical definitions and treatments of personal archives?

The dissertation's third chapter focuses on the question of how personal records have been defined and understood within the archival profession, and how this historical context impacts archivists working with personal digital records today. Beginning with early, foundational texts, archival definitions, attitudes, and practices relating to personal records will be explored. This chapter serves as the foundational literature review for the dissertation.

The following chapters comprise the primary research conducted. Each chapter explores collections that consist of or include personal digital records created within three overarching socio-technical environments. Each chapter discusses these collections through the dissertation's thematic lenses: Materiality; Custody and Control; and Privacy and Publicity. The fourth chapter examines personal records created, stored, and acquired by archivists on local storage devices, including removable storage media and personal computers. The fifth chapter examines personal digital records created, stored, and collected on the World Wide Web. These are collections that are accessible on the open web, including public-facing websites and blogs. The sixth chapter studies personal digital records that have been created within networked social platforms, including Twitter and other social media platforms.

The final chapter summarizes findings across all of these collections, reflecting on the approaches taken to working with them, and highlighting areas of concern and potential interest

for archivists who collect and care for personal digital records. Areas of future research will be identified as well.

3.0 What are Personal Archives?

As individuals in society, most of us are likely to, "consciously or not, create a personal archive that documents the different facets" of our activities, experiences, and relationships.[70] In spite of their ubiquity, defining personal records has proven a persistent challenge to the archival profession. Early foundational works in the field defined personal records largely in the negative – by describing how they failed to meet the criteria for archival status.[71] This exclusion can be read through the lens of broken world thinking: explanations of how and why personal records are not archives can shed light on the construction of archival status within the profession. Through a review of the literature from this area of archival studies, this chapter will explore the history of personal records in the archival field, and establish a definition of personal digital records that will inform the dissertation research that follows.

This chapter situates personal digital records within the broader literature on personal records in the archival profession. As outlined in the introduction, the archival collections examined throughout this dissertation are viewed through the lens of three primary themes: Materiality; Custody and Control; and Privacy and Publicity. Scholarship pertaining to these themes is discussed through the following sections of this literature review. With this framework in mind, this chapter begins with a discussion of early archival theory and the professional traditions that helped to shape collecting practices for personal records. Because this research is primarily concerned with personal records that are created and collected digitally, contemporary

70 Caroline Williams, "Personal Papers: Perceptions and Practices," in *What Are Archives? Cultural and Theoretical Perspectives: A Reader*, ed. Louise Craven.
71 Robert Fisher, In Search of a Theory of Private Archives."

definitions of personal records and the characteristics of personal digital records, specifically, will then be discussed. Ethical issues associated with personal archives will be explored in order to provide context for the ethical considerations and approaches taken in the cases discussed throughout the dissertation, and conceptual frameworks for resisting a binary approach to classifying records as personal or professional will be introduced. Finally, the chapter concludes with a description of personal records as an archival genre.

As this chapter will demonstrate, personal records are created in a wide range of formats and styles, for a multitude of purposes. Ultimately, the resistance of personal records to standardization and broad systems of classification - the same qualities that have made them so difficult to define - are in fact their defining characteristics.

3.1 Personal Records in Archival Theory

Early archival scholarship established a firm distinction between "archives" and "historical manuscripts," the former being the records of organizations and the latter the personal papers of individuals and families. As Riva Pollard has written, those personal papers, or historical manuscripts, have "been notably neglected in the course of the development of archival theory," a gap she has examined in the context of its impact on archival appraisal.[72] The effect of this neglect, Adrian Cunningham as observed, has been that archivists who acquire or care for personal records have been left to "puzzle over the relevance and applicability of archival theory to our particular

72 Riva Pollard, "The Appraisal of Personal Papers: A Critical Literature Review," *Archivaria* 52 (2001): 137.

circumstances."[73] Indeed, much contemporary archival scholarship continues to interrogate these very issues.

In order to more fully understand the contemporary literature on the subject of personal records, it is necessary to first discuss how and why these records have been excluded or, as Jennifer Douglas has described them, "sidelined" in the development of archival theory. Archival theory provides archivists with a framework for understanding the nature of records, in order to identify those properties of a record that merit long-term stewardship.[74] Theory provides a basis for methodological approaches in professional practice; by beginning with questions about the properties of archival materials, practical approaches may be more effectively designed to specifically suit those properties.[75] Because early and foundational developments in archival theory have excluded personal records, subsequent professional literature addressing their care has been lacking, both in quantity and in the provision of specific guidance.[76] And treatments of personal records have been inconsistent, largely due to the fact that their care has been influenced by two separate professional traditions, with markedly different theoretical underpinnings.

3.1.1 The Historical Manuscripts and Public Archives Traditions

The collection and care of personal records in the United States has been associated primarily with two professional paradigms: the historical manuscripts tradition and the public archives tradition. The historical manuscripts tradition, most commonly associated with personal

73 Adrian Cunningham, "Beyond the Pale?" 23.
74 Terry Eastwood, "What is Archival Theory and Why is it Important?" *Archivaria* 37 (1994): 125
75 Eastwood, "What is Archival Theory and Why is it Important?" 126
76 Pollard, "The Appraisal of Personal Papers,"140

records from the 18th century through the mid-20th century, relied upon techniques grounded in the principles of librarianship.[77] The public archives tradition, which rose to greater prominence in the 1960s, was informed by the archival profession in 19th century Europe.[78] Until the mid-20th century, personal records were generally collected and cared for first by private collectors and antiquarians, and later by libraries, museums, and historical societies.[79] Archives, conversely, primarily acquired and cared for records created and accumulated within organizational settings.[80]

Though records of informational and evidential value were acquired, preserved, and made accessible to researchers by institutions working within both of these traditions, there were significant differences in the approaches taken by each to accomplish those activities. Key among these was the fact that organizations operating within the historical manuscripts tradition employed theoretical and practical approaches from librarianship, including item-level description, subject-based organizational schemas, and collection development policies developed to serve the needs of researchers. Public archives, conversely, described records in the aggregate, organized them according to creator rather than subject, and received records from the organizational bodies they served, rather than building collections deliberately.

Practices within the public archives tradition were grounded in the principles of provenance and original order.[81] Provenance refers to the origin or source of records, but it also encompasses information *about* the "origins, custody, and ownership."[82] taking into consideration the chain of custody between the time when a body of records is created and the time at which it is acquired by

77 Richard C. Berner, *Archival Theory and Practice in the United States: A Historical Analysis.* (University of Washington Press, 1983), 1.

78 Berner, *Archival Theory and Practice in the United States: A Historical Analysis*, 1-2.

79 Gilliland-Swetland, "The Provenance of a Profession," 161.

80 Gilliland-Swetland, "The Provenance of a Profession."

81 Gilliland-Swetland, "The Provenance of a Profession."

82 "Provenance," *Dictionary of Archives Terminology*, https://dictionary.archivists.org/entry/provenance.html.

an archives. The principle of provenance incorporates the French concept of *respect des fonds*, which requires that records of unique provenance be kept separate from one another in order to preserve their original contexts and maintain a trustworthy chain of custody.[83] The principle of original order refers to the "organization and sequence of records established by the creator of the records."[84] The principles of the public archives tradition suggest that records should be maintained in the original arrangement established by their creators in order to maintain any existing relationships between records and the evidence that may be provided by those very relationships. These principles supported the preservation of the context and integrity of aggregated records.

In service of the goal of providing researchers with direct access to records of interest, professionals working in institutions that operated according to the principles of the historical manuscripts tradition would, in certain situations, assemble subject-based collections, with materials culled from many collections, received from many creators. These are referred to as artificial collections, and stand in contrast to organic collections, which, as described above, adhere to the principle of provenance and are grouped according to their creator, in the original order. Artificial collections, in contrast, are typically organized according to a particular subject or event, in a system consistent with library classification. Laura Coles has used the example of a local historical society collecting and aggregating records related to the construction of the local railway to describe artificial collections. In this example, records that document the railroad construction are pulled from larger, organic collections such as those "of a local construction company which helped build the train station or the papers of a citizen who worked on the railway."[85] As Cole has

83 "Provenance," *Dictionary of Archives Terminology*.
84 "Original Order," *Dictionary of Archives Terminology*, https://dictionary.archivists.org/entry/original-order.html.
85 Laura Coles, "Organizing Archival Material," in *A Manual for Small Archives* (Vancouver, BC: Archives Association of British Columbia, 1998): 31. https://aabc.ca/media/6069/manualforsmallarchives.pdf.

suggested, these records, now grouped according to subject matter rather than creator, offer researchers some information about the subject at hand, but lack the contextual information provided by original order and provenance. Crucially, the chain of custody of these records is broken. Artificial collections are largely viewed as remnants of the historical manuscripts tradition, though, as this dissertation will illustrate, collections of personal digital records today take a range of shapes and forms, collected and arranged according to a variety of principles. The collections in the following chapters will be acquired in some instances according to their creator, in others according to the commercial platform in which they were created, and in still others according to a particular topic or theme. The tendency toward building artificial collections has arguably resurfaced in the digital space, and in particular, in social media archiving.

Because personal records were widely considered to be the purview of librarians and manuscript curators, they were collected, arranged, and described according to the principles of the historical manuscripts tradition, rather than those of the public archives tradition. The following section discusses the reasons frequently cited for excluding personal records from the public archives tradition.

3.1.2 Motivations for Excluding Personal Records

Archivists frequently turn to the *Manual for the Arrangement and Description of Archives* by Muller, Feith, and Fruin (commonly referred to in the literature as the Dutch manual), and the works of Sir Hilary Jenkinson and Theodore Schellenberg to understand the foundations of

archival theory and practice.[86] These works share many common perspectives, including a steadfast belief that personal records do not, and should not, fall within the archivist's purview. However, as Robert Fisher has observed, in establishing definitions of archives that specifically exclude personal materials, each of these authors has actually provided their readers with a characterization of personal records that has persisted into the 21st century.[87] Though these works primarily define personal records in the negative - by emphasizing how they are not archives - these characterizations constitute a productive starting point for examining the nature of personal records.

In fact, these archival theorists did maintain that personal materials were valuable, and that they did, in many respects, resemble archives. Muller, Feith, and Fruin wrote that "even private individuals may have archives," though they qualified this statement by suggesting that in order to be worthy of inclusion in the archives, these records would belong to individuals who were self-employed – independent merchants, for example, who kept "journals, cash books, received copies of letters sent, etc." for professional purposes.[88] The private papers of public officials were another exception to the exclusion of personal records. These materials, the authors of the Dutch manual explained, could "throw greatly desired light on the contents of the archival collection," and thus should be retained, albeit in a separate section of the collection inventory.[89] Later, Schellenberg would acknowledge that, indeed "the manuscript holdings of libraries cannot be differentiated

86 Because the focus of this dissertation is on English language archives, and particularly archives in North America, the theoretical lens is likewise limited to those works that have been influential in the development of this particular segment of the archival profession.

87 Robert Fisher, "In Search of a Theory of Private Archives," 2.

88 Muller, Feith, and Fruin. *Manual for the Arrangement and Description of Archives,* 20.

89 If there is not a public library nearby, and removing the papers from the local vicinity is not desirable because researchers may wish to reference them, the authors suggest that "one may form near the archival collection a library, whose description, in small archival collections, may perhaps be published as a supplement to the archival inventory." Muller, Feith, and Fruin, *Manual for the Arrangement and Description of Archives*, 155.

from archives on the basis of their form, their authorship, or their value. They may have come from similar sources, institutional or private; and they may be equally valuable for researchers."[90] Despite their similarities in form and function, records unaffiliated with professional activities were distinguished from "true archives." archival records.

As Terry Cook has explained, the Dutch manual is primarily concerned with "government, public, or corporate archives and their orderly transfer to archival repositories to preserve their original order and classification.[91] Personal records, created beyond the bounds of official structures and mandates, are resistant to such orderly transfer. The authors of the Dutch manual insisted that personal records are best left in the custody of librarians or curators in cultural heritage or memory institutions, a distinction that served to underscore the separation between the historical manuscripts and public archives traditions for years to come. This perspective was later accepted and advanced by both Jenkinson and Schellenberg.

Indeed, a reason frequently given for this separation or exclusion of personal records was the perception that unlike organizational archives historical manuscripts were deemed to be created and maintained in a haphazard, random manner. This perception was based, of course, on a conviction that organizational records actually *were* created in a systemic, organized manner. Individual or family papers were, conversely, deemed a random "conglomeration of papers," which were "gathered together in the strangest manner and lack the organic bond of an archival collection."[92] Both Jenkinson and Schellenberg affirmed that the archivist's ability to establish an "unbroken chain of custody" – to ensure that they were "actually records of the office that offers

90 Theodore Schellenberg, *Modern Archives: Principles and Techniques*, (Chicago: Society of American Archivists, 2003): 18.

91 Cook, "What is Past is Prologue: A History of Archival Ideas since 1898 and Future Paradigm Shift," *Archivaria* 43 (1997): 21.

92 Muller, Feith, and Fruin, *Manual for the Arrangement and Description of Archives*, 20.

them" – was essential to establishing archival status.[93] The unstructured, seemingly random ways in which personal records were created and maintained by individuals and families could not, in their estimation, ensure that unbroken chain of custody.

In addition to the systematic (or unsystematic, in the case of personal records) manner in which records were created, the motivations of their creators for generating or accumulating them was a further point of contention. Archives, as Jenkinson defined them, were "documents which formed part of an official transaction and were preserved for official reference."[94] Because of their transactional nature, Jenkinson considered them to be unselfconscious and thus more impartial. Impartiality, in Jenkinson's usage, did not mean that the creators of records were completely free of bias, but rather that the reasons for which the records were created could reliably ensure that they were not created for the purposes of posterity, as Terry Eastwood has explained.[95] Nonetheless, the perception that personal records were created self-consciously, and perhaps even with posterity in mind, was seen to discredit their viability as impartial evidence. Schellenberg would later take up this line of thinking as well, defining archives as records created or accumulated in the "accomplishment of official business." Fisher has reflected on these arguments, noting that for early archival theorists, "the presence of the personal, the intrusion of the self, compromised the impartiality of the record; no archivist could guarantee the impartiality of a personal narrative written with regard to the future."[96]

The motivations for collecting and preserving records constituted another difference between the public archives and historical manuscripts traditions. Jenkinson took a firm stance

93 Schellenberg, *Modern Archives*, 14.
94 Sir Hilary Jenkinson, *A Manual of Archive Administration Including the Problems of War Archives and Archive Making*, (London; Oxford; New York: The Clarendon Press, 1922): 21.
95 Eastwood, "What is Archival Theory and Why Does It Matter?" 127
96 Fisher, "In Search of a Theory of Private Archives," 9.

against active collecting, declaring that "Archives are not collected: I wish the word 'Collection' could be banished from the Archivist's vocabulary, if only to establish that important fact."[97] Seeking out and collecting aggregated records to satisfy researcher needs was a practice associated with librarianship and the historical manuscripts tradition, rather than the archival profession. Schellenberg later addressed the importance of understanding motivations for retaining archives in his own work. "To be archives," he wrote, "materials must be preserved for reasons other than those for which they were created or accumulated. These reasons may be both official and cultural ones."[98] However, he clarified, "their cultural values are incidental," rather than central.[99] Though they could not be counted as archives, Schellenberg did acknowledge the historical or cultural value of personal records, even citing the work of librarian Phyllis Mander Jones, who suggested that within personal accounts, one "finds a more personal contact with his subject, perhaps because private papers are more likely to reflect natural human prejudices and feelings."[100] While these "natural human prejudices and feelings" might be of interest to the user, they simultaneously embodied the failure of personal records to provide the impartial, objective evidence of activities found within archives.

Reviewing the discussion of personal records in these foundational texts, it becomes clear that personal records are not excluded on the basis of their evidential or informational value, but rather because they are not so readily compliant with the principles, standards, and procedures advanced for the professionalization of archival work. As Muller, Feith, and Fruin claimed, it was

97 Jenkinson, *The English Archivist: A New Profession* (An inaugural lecture for a new course in archive administration delivered at University College, London, October 14, 1947: London, 1948). Quoted in Schellenberg, *Modern Archives*, 19.

98 Schellenberg, *Modern Archives*, 13.

99 Schellenberg, *Modern Archives*, 17.

100 Phyllis Mander Jones, "Archivists and the Association," *The Australian Library Journal* 1, no. 4 (1952): 79. Quoted in Schellenberg, *Modern Archives*, 18.

to the advantage of the archivist to remove personal records from archival collections or, better, to refuse them altogether in the first place. Though they claim archival theory as the justification for this choice, they are more elaborate in their reasoning when discussing the impact on practice:

> "Theory demands this; for they do not satisfy our definition of an archival collection. But practice also requires it; for the archivist who has arranged his collection in conformity with our method according to the various administrative branches will naturally of his own accord keep out these documents which cannot fit into any of the headings adopted. He is embarrassed by them, since these documents interfere with good order in his depository. It is therefore absolutely necessary to exclude them from the collection proper."[101]

These foundational texts suggest that personal records should be excluded from the archival purview not because of they are personal, but because their idiosyncratic, "haphazard" characteristics disrupted the protocols established with the profession. Ultimately, as the archival profession further developed, this would be insufficient cause for exclusion, and yet personal records would maintain, to varying degrees, a status as "other" within the field. The following section outlines several of the developments that brought personal records under the umbrella of archives.

3.1.3 Converging Traditions: Bringing Personal Records into the Archives

By the mid-20th century, the historical manuscripts tradition began to lose prominence, with the public archives tradition gaining more traction not only in public archives, but in historical societies and other cultural heritage institutions. One reason for this was the deluge of incoming records - both personal and organizational - in the years after World War II, during which archivists

[101] Muller, Feith, and Fruin, *Manual for the Arrangement and Description of Archives*, 21.

and manuscript curators found themselves overwhelmed with both new acquisitions and existing processing backlogs.[102] Luke Gilliland-Swetland has suggested that by the 1950s, an apparent consensus had formed between manuscript curators, who agreed that principles from the public archives tradition, including aggregate description, original order, and Schellenberg's efficient and structured approach to appraisal, would allow them to more expediently process collections and make them accessible to users than the item-level description and laborious subject-based organization of librarianship.[103] This development resulted in a body of shared theoretical influences and practical applications among archival institutions and other cultural heritage institutions with manuscript collections. This shared body of knowledge was further cemented, as Gilliland-Swetland has argued, by another 20th century development: the movement of archival training from history departments to professional training programs in library and information science.[104] While few programs have offered dedicated courses in personal archives, as Jennifer Douglas has written, this confluence of training programs nonetheless brought together training that would encompass both professional traditions, making space for personal records within the archival field.[105]

Acceptance of personal records was further driven by the growing prominence of university and university archives within the archival profession. Academic archivists had adopted from the historical manuscripts tradition the practice of actively collecting records in accordance with the disciplinary areas for which the university was known and the interests of its research

102 Gilliland-Swetland, "The Provenance of a Profession,"162.

103 Gilliland-Swetland, "The Provenance of a Profession,"162.

104 Gilliland-Swetland, "The Provenance of a Profession,"169.

105 Jennifer Douglas has argued that just as personal records have been "sidelined in archival theory," so too have they been "sidelined in archival education." Douglas, "Getting Personal: Personal Archives in Archival Programs and Curricula" *Education for Information* 33, no. 2 (2017): 91.

communities.[106] Academic archives became a dominant force in professional organizations, such as the Society of American Archivists. Their increased stature and presence resulted in the increased influence and adoption of their practices and principles throughout the field. This included the strategy of building collections that included not only the records of their university and its faculty, but of prominent figures who would attract widespread research interest.

Even as personal records occupied a more secure place within the archival field, they have remained something of a niche interest, with a relatively small but dedicated number of archivists working to address the "silences concerning personal archives in mainstream archival theory."[107] Douglas has observed that, particularly since the 1990s, "there have been several small 'flurries' of activity related to personal archives."[108] These include special issues of journals, including *Archivaria*; single-authored monographs and volumes of collected essays; panel discussions at professional conferences; and even small annual conferences, such as Personal Digital Archiving.[109] Discussion in these arenas frequently considers the impact of the historical exclusion of personal records from archival theory and instruction and seeks to establish paths forward for archivists who care for personal archives. The following section draws upon this scholarship to consider how personal records, and personal digital records, are defined and understood within the field today.

106 Gilliland-Swetland, "The Provenance of a Profession," 164.

107 Catherine Hobbs, "The Character of Personal Archives: Reflections on the Value of Records of Individuals," *Archivaria* 52 (2001): 125.

108 Douglas, "Getting Personal," 92.

109 See, for example, *Archivaria* 76 (Fall 2013), https://archivaria.ca/index.php/archivaria/issue/view/459; Richard J. Cox, *Personal Archives and a New Archival Calling;* Christopher A. Lee, *I, Digital: Personal Collections in the Digital Era*; Personal Digital Archiving 2018, https://sites.lib.uh.edu/pda18/.

3.2 Defining (and Redefining) Personal Records Today

Though it is now quite commonplace to find personal records within archives of all kinds, the question of how, precisely, to define them is one that remains, in Caroline Williams' words, "vexed."[110] While the works that established the foundation of the modern archival profession largely defined personal records in the negative – by emphasizing how they failed to meet the requirements of true archives – modern archival scholarship has endeavored to take a more direct approach to explicating personal records.

The Society of American Archivists (SAA) *Dictionary of Archives Terminology* defines a record as, "data or information that has been fixed on some medium; that has content, context, and structure; and that is used as an extension of human memory or to demonstrate accountability," or as "data or information in a fixed form that is created or received in the course of individual or institutional activity and set aside (preserved) as evidence of that activity for future reference."[111] The "content, context, and structure" triad is one that appears frequently within the archival literature. Content indicates the "text, data, symbols, numerals, images, sound, graphics, and other information that make up the substance of the record."[112] Context refers to the "organizational, functional, and operational circumstances surrounding a record's creation, receipt, storage, or use."[113] Structure indicates "a record's physical characteristics and internal organization of the content."[114] Content and structure can be thought of as internal to the record, used to conceptualize and describe a record itself, while context establishes its relationship to other records and the

110 Caroline Williams, "Personal Papers: Perceptions and Practices," 57.
111 "Record," *Dictionary of Archives Terminology,* https://dictionary.archivists.org/entry/record.html.
112 "Record," *Dictionary of Archives Terminology*.
113 "Record," *Dictionary of Archives Terminology*.
114 "Record," *Dictionary of Archives Terminology*.

environment in which it was created, and is thus external to the record itself.[115] This triad provides a useful template for articulating the various aspects of a record that archivists may wish to preserve. This definition of a record bears a resemblance to the Documentalist definition of documents that provides a basis for data selection in this dissertation. Briet defines a document as "any concrete or indexical sign, preserved or recorded toward the ends of representing, of reconstituting, or of proving a physical or intellectual phenomenon."[116] Much like the definition provided by the SAA, the definition outlined in the Documentation movement explains that records have materiality; they are created intentionally; they are processed or turned into documents; and they are understood or perceived as documents by people.[117] This understanding of a record is well-suited to personal records, in part because it acknowledges the essential role of individuals or groups of people in both the creation of the record itself and the processes that bestow different kinds of value (such as archival value) on the record after its creation. Drawing upon this understanding of what constitutes a record, the remainder of this section explores what it means for a record to be personal.

A multitude of definitions for personal records or personal archives can be found in the archival literature, and they vary considerably in both scope and perspective. Traces of the characterizations provided in early theoretical texts can be detected in many of these modern attempts at definition development. The SAA's *Dictionary of Archival Terminology*, for example, advances the tradition of situating personal records alongside or in contrast to organizational or professional records, defining them as "documents in any format that provide evidence of an

115 Luciana Duranti and Kenneth Thibodeau, "The Concept of Record in Interactive, Experiential and Dynamic Environments: The View of InterPARES." *Archival Science* 6, no. 1 (2006): 29.

116 Suzanne Briet, *What is Documentation?* Translated by Ronald E. Day and Laurent Martinet (Lanham, MD: The Scarecrow Press, Inc., 2006), 10.

117 Buckland, "What is a "Document?" 806.

individual's activities," or, alternately, "unofficial documents in any format kept by an individual at a place of work."[118] The ongoing distinction between personal records and organizational records is, in some instances, even more overt, as in Gregory Hunter's assertion that "archives are generated by organizations or institutions; manuscripts are generated by individuals or families," and "the custodian of organizational records is called an archivist, while the custodian of personal papers is called a manuscript curator."[119] Other archival scholars, including Richard Cox, have pushed back against this distinction, arguing that "there really is no room for disagreement. Archivists are archivists. Archives are archives. Archives are composed of records. Historical manuscripts are composed of records, and they constitute archives. Manuscript curators are responsible for records *and* archives."[120] Other definitions fall somewhere between these stances, while generally accepting the proposition that personal records are indeed archives. Williams has suggested that "personally created papers may certainly exhibit the attributes required of records. They provide recorded evidence of the activities of the creator, whether official (letter to the bank) or personal (text message to son)."[121] Though personal records are now largely accepted as archival records, the influence of their historical exclusion is still evident in the professional literature.

One aspect of this legacy is the quality of standardization, or lack thereof, in personal records. Traces of Schellenberg's claim that "while archives grow out of some regular functional activity, historical manuscripts, in contrast, are usually the product of a spontaneous expression of thought or feeling," as well as Muller, Feith, and Fruin's earlier claim that these materials have

118 "Personal Papers," *Dictionary of Archival Terminology*, https://dictionary.archivists.org/entry/personal-papers.html.

119 Gregory S. Hunter, *Developing and Maintaining Practical Archives* (New York: Neil-Schumann, 2004):

120 Richard Cox, *Managing Records as Evidence and Information*, (Westport, CT: Quorum, 2001), paraphrased in Caroline Williams, *Managing Archives: Foundations, Principles, and Practice*, (Oxford: Chandos Publishing, 2006).

121 Williams, "Personal Papers: Perceptions and Practices," 59.

generally been "gathered together in the strangest manner and lack the organic bond of an archival collection" can be observed in these contemporary descriptions of personal records.[122] Amber Cushing has addressed this line of thinking, writing that "while public records are often systematic, personal collections are ambiguous and vary from collection to collection: there is no standard format for each individual record."[123] In her definition of personal archives, Catherine Hobbs has synthesized some of these traits, incorporating consideration of the many ways individuals create records and maintain them over time for future use or reference:

> "Personal archives are formed because of the needs, desires, and predilections of their creators to create and keep documents (not for an administrative purpose or because of a legal requirement). Personal archives are controlled entirely by private individuals before they enter a repository. Because individuals create documentation for personal reasons outside an administrative context, they dictate the forms documents take, the genres of their writing, and the changes made during their use."[124]

In Hobbs' definition, the spontaneous, nonstandard nature of personal recordkeeping is not a liability or a fault, but merely a characteristic to be acknowledged. Further, in this formulation the idiosyncratic approach to documentation, because it is not informed by an administrative environment, has the potential to tell us more about the creator or the context of creation.

Though foundational archival texts advised against active collecting, and building collections in response to, or anticipation of, researcher interests, collective memory and documentary heritage have emerged as central concerns in the preservation of personal archives. Sue McKemmish has been credited as one of the first archival scholars to directly address the

122 Schellenberg, *Modern Archives*, 18; Muller, Feith, and Fruin. *Manual for the Arrangement and Description of Archives*, 20.

123 Cushing, "Highlighting the Archives Perspective in the Personal Digital Archiving Discussion," *Library Hi* Tech 28, no. 2 (2010): 302.

124 Hobbs, "Reenvisioning the Personal: Reframing Traces of Individual Life," 213.

question of how collections of personal records "might constitute 'enduring value' to a society."[125] In "Evidence of Me," McKemmish argued that the practice of personal recordkeeping was an act of "evidencing and memorializing" the activities, experiences, and relationships with others - a way of placing one's lived experience within a broader context.[126] The archivist's task, then, is to collect and preserve that evidence, and to ensure that it remains an accessible part of a society's memory and cultural heritage.[127] In practice, the appraisal and acquisition of personal records has proven to be uneven. Graeme Powell has suggested that because the acquisition of personal records "has been driven strongly by the needs of researchers," the lived experiences conveyed by archives have been unbalanced, and "many groups in society, both past and present, are represented in only the most meager way."[128] Indeed, statements like Mary Lynn McCree's suggestion that the collecting archivists "primary responsibility is to create a focused body of materials that informs the scholar," are not uncommon in collection development literature.[129] While this strategy fulfills a mission to serve users, it implicitly values the researcher over the creator. Powell has observed, for example, that personal papers in archives tend most frequently to be those of political figures and writers, in part because "political and literary historians have always been conspicuous in reading rooms and have worked closely with librarians and archivists."[130] In this respect, the promise of personal archives to provide a diverse or representative "evidence of us," is not entirely

125 Pollard, "The Appraisal of Personal Papers," 146.

126 McKemmish, "Evidence of Me," 29.

127 McKemmish, "Evidence of Me," 30

128 Graeme Powell, "The Collecting of Personal and Private Papers in Australia," *Archives & Manuscripts* 24, no. 1 (1996): 72

129 Mary Lynn McCree, "Good Sense and Good Judgment: Defining Collections and Collecting," in *A Modern Archives Reader: Basic Readings on Theory and Practice*, ed. Maygene F. Daniels and Timothy Walsh (Washington DC: National Archives Trust Board, 1984): 105

130 Powell, "The Collecting of Personal and Private Papers in Australia, 73. Powell's observation resonates with this dissertation research. Many of the collections examined in the following chapters are the personal archives of writers and other prominent cultural figures.

fulfilled. As Williams has suggested, "the survival of the mass of personal papers generated by individuals of lesser public standing… is less likely to be assured and has always been to some degree serendipitous," with occasional instances of such records being collected within local historical societies or archives with collecting focuses on particular occupations or geographic areas.[131]

Finally, while personal archives are most frequently described as those created *by* an individual, there has been growing recognition within the field of the fact that personal records are also created *about* them as well. They might be created by other individuals who know the person in question, but they might also be created by official agencies. Christopher A. Lee has defined the personal archive as an "aggregate of an individual's personal traces that the individual *or someone else* has identified and attempts to manage over time as a relatively coherent unit in order to reflect something important about that individual" [emphasis added].[132] Personal records of this variety are "generated and maintained *by* organizations - schools, hospitals, and the state," as Williams has likewise suggested.[133] They are typically categorized by archivists as official or organizational records, rather than personal records, "because they are generated by organizations with a specific business function and purpose. Some official personal records are created by official bodies and retained by individuals for evidentiary purposes: passports, drivers licenses, and examination certificates."[134] However, these records can be simultaneously organization and

131 Williams, "Personal Papers: Perceptions and Practices," 55.

132 Christopher A. Lee, "Introduction," *I, Digital: Personal Collections in the Digital Era*, ed. Christopher A Lee. (Chicago: Society of American Archivists, 2011): 3.

133 Williams, "Personal Papers: Perceptions and Practices," 56.

134 Williams, "Personal Papers: Perceptions and Practices," 56.

personal: a passport, for example, can over time serve as a keepsake travel journal for its bearer, in addition to a legal document required for international travel.[135]

3.3 Personal Digital Archives

If personal records have not historically found equal footing with organizational records in archival theory and practice, the proliferation of born-digital personal records has helped to bring them to the forefront. Neil Beagrie has referred to creators of digital and web-based personal records "Generation C," with "C" standing for "content." The "Generation C phenomenon, he wrote, referred to "a perceptible consumer shift from consumption to personal creation, customization, and co-production of digital content."[136] As Beagrie suggested, the ability to easily and affordably create and store an abundance of digital records locally and on the World Wide Web has resulted in a preponderance of digital records to be addressed by archivists. Additionally, as Williams has suggested, in a digital environment, there may be "an increasing conscious desire to *create* and *collect* personal history too, based on the aspiration to leave some kind of a footprint or trace behind."[137] While diaries, correspondence, and memoirs have always afforded this opportunity, technological developments provide a greater number of individuals with the tools to document their own lives and experiences, so much so that they might support the transformation of documentary heritage:

135 Frameworks for thinking beyond traditional notions of "official" or personal" will be discussed in greater detail in the section "Reimagining Personal and Organizational Classification" below.
136 Beagrie, "Plenty of Room at the Bottom?"
137 Williams, "Personal Papers: Perceptions and Practices," 56.

> "After all anyone can start a weblog and web-based software is the medium for many a community archive. Minority and indigenous groups, community archives, and individuals are creating new spaces - elbowing into the space traditionally perceived as inhabited by conventional archives, whether organizational or personal. Records created in this way may be more self-conscious, and may be generated in order to actively document current actions and concerns, rather than be simply the unconscious residue of past activities and transactions."[138]

With digital and web-based tools, individuals have a host of new ways to not only create personal records, but to disseminate them to their audiences. This ability has led to new opportunities to collect and preserve an archival record that includes the documentation of everyday life. At the same time, the digital environment has introduced new challenges – technological, legal, and ethical – for the archivists who collect and preserve them.

Expanding upon the literature dedicated to the subject of personal archives is a body of work focused specifically on personal digital archives, and the nascent area of personal digital archiv*ing*. Personal digital records are the records created by individuals in digital environments and formats; in the past several decades, many people have come to create a significant percentage of their personal records – including correspondence, photographs and videos, and personal reflections and journals – in digital form.[139] As many have suggested, personal digital records have much in common with their non-digital predecessors, particularly their ability to provide evidence of an individual's lived experience. However, born-digital personal records differ from paper-based personal records in meaningful ways.

Among these is in the resistance of digital records to the shoebox metaphor, a popular strategy for describing personal archives. As Catherine Marshall has explained, the shoebox

138 Williams, "Personal Papers: Perceptions and Practices," 56.
139 Toby Burrows, "Personal Electronic Archives: Collecting the Digital Me," *OCLS Systems and Services: International Digital Library Perspectives* 22, no. 2 (2006): 85-86.

metaphor refers to an individual's collection of valuable documents and keepsakes, which might be stored in a box and tucked away under a bed, on a shelf, or in an attic.[140] This metaphor breaks down in the digital sphere for a variety of reasons, with two that stand out as particularly important. First, the shoebox model proposes benign neglect as a kind of "accidental" preservation model; by putting the materials aside and accessing them infrequently, they are saved from the wear and tear of frequent use. Digital records, however, require regular maintenance, including software updates, translation, and migration, in order to persist over time.[141] Much has been written about the technological challenges associated with preserving born-digital records. Bernadette Houghton cites the short lifespans of many hardware and software systems, which are especially short when compared with those of paper-based records.[142] Williams notes the problem of receiving records in obsolete formats that cannot be read with contemporary software, rendering their contents inaccessible.[143] Migration and emulation, two frequently employed digital preservation strategies, each required both technological expertise and infrastructure.[144]

Additionally, digital files, whether created on a personal computer or the World Wide Web, are decentralized, and resistant to the type of tidy storage depicted in the shoebox metaphor.[145] Cushing has written that, "a personal digital archive refers to digital items within an individual's control that have been stored and maintained by the individual."[146] This statement holds true

140 Catherine C. Marshall, "Rethinking Personal Digital Archiving Part 1: Four Challenges from the Field," *D-Lib Magazine* 14, no. 3 (2008), http://www.dlib.org/dlib/march08/marshall/03marshall-pt1.html.

141 Rothenberg, "Ensuring the Longevity of Digital Information," 11.

142 Bernadette Houghton, "Preservation Challenges in the Digital Age," *D-Lib Magazine* 22, no. 7 (2016), http://www.dlib.org/dlib/july16/houghton/07houghton.html.

143 Williams, "Personal Papers: Perceptions and Practices," 65.

144 Kenneth Thibodeau, "Overview of Technological Approaches to Digital Preservation and Challenges in the Coming Years," Council on Library and Information Resources Reports (2002), https://www.clir.org/pubs/reports/pub107/thibodeau/.

145 Marshall, "Rethinking Personal Digital Archiving Part 1."

146 Cushing, "Highlighting the Archives Perspective in the Personal Digital Archiving Discussion," 301.

primarily when applied to those records that have been created and stored on a local storage device, though even within a single device records might be easily displaced, deleted, or forgotten about. In a networked environment, individual record creators share control with, and arguably cede it to, the commercial platforms in which records are created and stored. These records call to mind Amelia Acker's conception of "born-networked" records. Writing specifically about records created with mobile information communication technologies, Acker has explained that these records are born-networked "because when they are created and transmitted, they become subject to a host of network architecture, standards, machines, fiber, and wires" that support their transmission.[147] The decentralized, networked infrastructure supporting personal digital records complicates storage, access, and control for personal archives. However, "this reality," as Cushing has suggested, "provides archivists the opportunity to enter discussions about the concept of place in association with digital records and distributed storage in order to add to the greater conversation about the personal digital archiving of individuals."[148] Within the shoebox metaphor, a record creator might safely store their materials in a box for decades before safely depositing them in an archives; in the digital environment, more proactive measures have been deemed necessary.

3.3.1 Personal Digital Archiving

The term "personal digital archiving" (PDA) refers to the collection, management, and preservation of individual and family records that are created in digital form.[149] This field is closely

147 Amelia Acker, "Radical Appraisal Practices and the Mobile Forensic Imaginary," *Archive Journal* (2015), http://www.archivejournal.net/?p=6204.

148 Cushing, "Highlighting the Archives Perspective in the Personal Digital Archiving Discussion," 306.

149 Gabriela Redwine, "Personal Digital Archiving," Digital Preservation Coalition Technology Watch Report 15-01 December 2015, https://www.dpconline.org/docs/technology-watch-reports/1460-twr15-01/file.

connected to that of personal information management (PIM), which explores the ways in which individuals identify, organize, and access their personal data.[150] Personal digital archiving typically addresses the actions taken by individuals to organize and preserve their *own* records, though archivists and other information professionals can play a number of vital roles in this realm.[151] Specifically, archivists engaged in this area offer primarily and education, supporting their constituents in the preservation of their personal records rather than assuming custody of those records within archives.[152]

Archivists engaged in supporting the personal digital archiving practices of the public operate, in part, according to Cunningham's directive "that archivists should be present in the pre-custodial (before items are in the custody of the archivist) phase of the records life cycle instead of confined to the inactive stage."[153] While this suggestion was originally made with collections that would eventually be acquired by an institution in mind, this perspective has been taken up within the profession and expanded to include the support of records that may not ever be located within an archival repository. Here, the archivist's role is to make recordkeeping expertise accessible to members of the public as they develop and care for their own personal digital records.[154] To support the non-custodial preservation of personal records, archivists might share their expertise by teaching individuals how to create records in stable formats, download their social media archives, or organize and store their personal digital records.[155] They might also

150 Vanessa Reyes, "We Created It, Now How Do We Save It? Issues in Preserving Personal Information, A Review," *Preservation, Digital Technology, & Culture* 42, no. 3 (2013): 150.
151 Reyes "We Created It, Now How Do We Save It?" 153.
152 Cox, *Personal Archives and a New Archival Calling*, vii.
153 Cunningham 1994, p. 101
154 Cushing, "Highlighting the Archives Perspective in the Personal Digital Archiving Discussion," 304
155 See Jamie Wittenberg and Celia Emmelhainz, "Access, Annotate, Export" and Melody Condron, "Archiving Social Media," in Brianna Marshall ed., *The Complete Guide to Personal Digital Archiving* (Chicago: ALA Editions, 2017).

provide hardware and software to support access to obsolete storage media or file formats.[156] Educational initiatives and physical spaces dedicated to these activities can be found in many institutions, including archives, academic and public libraries, and museums. While the focus of this dissertation is on personal digital archives that have been collected and preserved within memory institutions, it is important to acknowledge that the archival profession has also taken steps to support personal digital archives, and to engage their communities, beyond their own collection development.

3.4 Reimagining Personal and Organizational Categorizations

In practice, the distinction between those records that are considered personal and those that are considered professional is not as straightforward as foundational archival texts have suggested. Often, it can be "difficult to discern the precise boundary between public, official, and personal."[157] In his deconstructive reading of McKemmish's "Evidence of Me," Verne Harris argued that in fact, the boundaries between personal and professional recordkeeping spaces "is far from untroubled."[158] As Harris suggested, the separation between the personal and the public is one that "every individual makes, but it is determined by an indeterminable and shifting context of cultural and societal layerings."[159] Given this, how can fixed or rigid classifications such as personal or professional be adequately or effectively applied? These categories reflect professional

156 Memory Lab Network, https://memorylabnetwork.github.io/.
157 Williams, "Personal Papers: Perceptions and Practices," 55
158 Verne Harris, "On the Back of a Tiger: Deconstructive Possibilities in 'Evidence of Me,'" *Archives & Manuscripts* 29 no. 1 (2001): 18.
159 Harris, "On the Back of a Tiger," 19.

archival notions of order, rather than the lived experience of recordkeeping. In attempting to account for this complicated state of affairs, a number of archivists have developed additional theoretical frameworks for acknowledging the permeable boundaries between the various spheres of life in which record creation occurs.

While processing archival collections of collocated corporate and personal records from family businesses in Nova Scotia, Creighton Barrett developed the concept of the "work identity of the creator(s)" to provide context for instances in which personal and family records were collocated within corporate fonds.[160] Barrett suggested that the physical and intellectual intermingling of personal records with professional is often not incidental, but is instead an intentional and meaningful act on the part of the creators. As such, it should be understood and arranged as a holistic form of recordkeeping, rather than an unfortunate combination of two discrete collections. If there are countless factors that have the potential to influence the formation of personal identity, then by extension there are as many factors that inform the shape and nature of personal archives; Barrett identified professional activity as one such significant factor.[161] Within this framework, an individual with a strong work identity might establish an archive in which the personal and professional are deeply intertwined, reflecting the strong role that professional activities play in the development of their concept of self; conversely, an individual with a lesser degree of work identity might attempt to keep professional activities and records more removed from their life outside of work.[162]

160 Creighton Barrett, "Respect Which Fonds? Personal Archives and Family Businesses in Nova Scotia," *Archivaria* 76 (2013): 86.

161 Barrett, "Respect Which Fonds?" 87.

162 Barrett, "Respect Which Fonds?" 87.

Another model for rethinking a "personal vs. professional" binary comes from Neil Beagrie, who has proposed the "public persona" a concept that highlights the intersections between the personal and the professional by considering the degrees to which records are visible and accessible to their various publics.[163] Beagrie suggested that personal digital archives are often "composites drawing materials from an individual's private life, work, and education, as well as from external communities and content sources."[164] The public persona exists at the intersection of one's private persona, work persona, and those external communities and content (Figure 1).

Figure 1 Beagrie's Public Persona

Barrett's "work identity" and Beagrie's "public persona" are useful concepts for the reconsideration of the various points of intersection between the personal and the professional, and

163 Beagrie, "Plenty of Room at the Bottom?"
164 Beagrie, "Plenty of Room at the Bottom?

the potential for records from both of these spheres to bear witness to a single life. However, both of these formulations nonetheless continue to rely upon the existence of identifiable personal and professional activities and records. Other theoretical frameworks have attempted to transcend those distinctions might more accurately reflect the complexity of the lives and activities recorded in one's archives.

One such framework is that of the "archival multiverse," advanced by Sue McKemmish and Michael Piggott. The archival multiverse is proposed as an alternative to the "binary opposition of the personal and corporate archive - an either/or view of the archival world."[165] McKemmish and Piggott argue that by working directly with record creators and learning about the societal contexts in which recordkeeping takes place, archivists can see more clearly the ways in which "organizational" and "personal" are categorizes imposed by the archival profession, rather than organic qualities inherent to records or the activities that produce them.[166] Their conception of the archival multiverse convincingly argues for an approach to archival work that resists standardization and tidy classification, and instead acknowledges the myriad characteristics and traits that any record or collection might embody at different points in time, or for different users.

The concept of "activation," advanced by Eric Ketelaar, provides another lens through which archivists might rethink traditional notions of personal and professional.[167] Ketelaar suggests that how a record is defined or understood ultimately has less to do with the record itself, or even the purposes for which it was created, and much more to do with the ways in which any user

165 Sue McKemmish and Michael Piggott, "Toward the Archival Multiverse: Challenging the Binary Opposition of the Personal and Corporate Archive in Modern Archival Theory and Practice," *Archivaria* 76 (2013): 111–14.
166 McKemmish and Piggott, "Toward the Archival Multiverse," 113.
167 Eric Ketelaar, "Tacit Narratives: The Meaning of Archives," *Archival Science* 12, no. 2 (2012). 131-141.

encounters or engages with it; "every interaction, intervention, interrogation, and interpretation by creator, user, and archivist is an activation of the record.".[168] Jennifer Douglas and Allison Mills have argued for activation as an effective way of conceptualizing personal records in particular.[169] Rather than assign records to the category of personal or organizational, they suggest, archivists might instead focus on how records are experienced or activated by a more diverse body of stakeholders. As in the theoretical framing of the archival multiverse, activation allows a single record to occupy many roles. Scholarship on the case records of care-leavers offers further insight into what these frameworks might look like in practice.[170] Care-leavers are individuals who were raised, partially or entirely, within institutions run by religious, charitable or government organizations, or under the notional guardianship of the state as state wards."[171] Under the Freedom of Information Act, Care-leavers are able to access the case files and other records created by the institutions charged with their care, an experience that Jacqueline Z. Wilson and Frank Golding have explored in their research. Archival theory would suggest that these records are organizational: records created in the process of conducting official activities. However, as Wilson and Golding have argued, these records reveal much about the personal opinions and emotional responses of their creators through the range in their content and tone, encompassing both "dry, bureaucratic 'objectivity,'" and "highly personal judgmental commentary."[172] The affective experiences of Care-leavers upon accessing their records are well documented, encompassing both

168 Eric Ketelaar, "Tacit Narratives: The Meaning of Archives," *Archival Science* 12, no. 2 (2012): 131.

169 Jennifer Douglas and Allison Mills, "From the Sidelines to the Center: Reconsidering the Potential of the Personal in Archives," *Archival Science* 18, no. 3 (2018): 257-277.

170 Jacqueline Z. Wilson and Frank Golding, "Latent Scrutiny: Personal Archives as Perpetual Mementos of the Official Gaze," *Archival Science* 16, no. 1 (2016): 93-109.

171 Wilson and Golding, "Latent Scrutiny," 94.

172 Wilson and Golding, "Latent Scrutiny," 96.

trauma and vindication.[173] These records stand as an example of the personal experiences and emotions that can be embodied in both the creation of so-called official records and the experience of accessing and reading them. Indeed, as William and Golding have suggested, these records offer us much to consider now, when the archival profession stands "at a moment of radical re-evaluation of the nature and significance of personal records."[174]

3.5 Ethical Issues in Personal Archives

Given the affective potential of personal records – broadly defined – archival ethics are central to this work. Among the myriad ethical issues pertinent to personal records, this dissertation is primarily concerned with the ethical dimensions of record acquisition, privacy, and access and use. These areas recall three of the areas of information ethics identified by Richard O. Mason: property or ownership, privacy, and accessibility.[175] While these areas have always been relevant to issues in personal archives, Luciana Duranti and Corinne Rogers have argued that they are "especially pressing" in the digital and cloud-based environments in which so many personal records are now created.[176] Literature related to these issues is reviewed in this section, and explored throughout the remaining chapters in the context of specific archival collections.

173 Wilson and Golding, "Latent Scrutiny," 107.

174 Wilson and Golding, "Latent Scrutiny," 107.

175 Richard O. Mason, "Four Ethical Issues of the Information Age," *MIS Quarterly* 10, no. 1 (1986): 5. Mason's fourth area, accuracy, while significant within the broader field of information ethics, is beyond the scope of this dissertation, which does not address issues of accuracy, misinformation, or disinformation.

176 Corinne Rogers and Luciana Duranti, "Ethics in the Cloud," *Journal of Contemporary Archival Studies* 4, no. 2 (2017): 6.

The Society of American Archivists addresses issues related to appraisal and selection as well as archival custody in its Core Values Statement. The section on Selection addresses the vital role of appraisal in archival work, stating that, in recognizing that "because of the cost of long-term retention and the challenges of accessibility most of the documents and records created in modern society cannot be kept."[177] This being the case, archivists "acknowledge and accept the responsibility of serving as active agents in shaping and interpreting the documentation of the past."[178] This is a significant departure from the notion of the archivist as a neutral custodian of records.

While privacy encompasses both legal and ethical concerns within archival science, this dissertation is focused primarily on the ethical dimensions of privacy. It can, however, be difficult to untangle these two approaches to thinking about privacy. The Society of American Archivists' Code of Ethics for Archivists begins its section on privacy by stating that "archivists recognize that privacy is sanctioned by law."[179] However, a sensitive treatment of personal privacy within archival collections or research may lend itself to more restrictive or conservative measures than are required by law. The Code of Ethics further elaborates that "archivists place access restrictions on collections to ensure that privacy and confidentiality are maintained, particularly for individuals or groups who have no voice or role in collections' creation, retention, or public use." In addition, the Code states that "archivists promote the respectful use of culturally sensitive materials in their

177 "Core Values of Archivists," Society of American Archivists, https://www2.archivists.org/statements/saa-core-values-statement-and-code-of-ethics.
178 "Core Values of Archivists," Society of American Archivists.
179 "Code of Ethics for Archivists," Society of American Archivists, https://www2.archivists.org/statements/saa-core-values-statement-and-code-of-ethics.

care by encouraging researchers to consult with communities of origin, recognizing that privacy has both legal and cultural dimensions."[180]

This dissertation takes Nissenbaum's work on privacy as contextual integrity as its primary lens for considering how personal privacy can be both protected and threatened through archival work. This framework suggests that individuals create and use records in specific contexts. Their privacy is violated when personal records, or personal information, are passed from one context to another without their consent.[181] The collections explored in the following chapters will include both those in which the creator chose to deposit their records in the archives, and those in which records were collected without the creator's knowledge or consent. Danielson has argued that "the violation of privacy is an intrinsic and unavoidable part of archival work, because it involves the secondary use of documents, which were created for another, so called primary, purpose."[182] In all personal collections, risks to privacy are present and cannot be avoided, but might be mitigated by degrees.

In order to protect both record creators and subjects, it has long been common practice to place restrictions on personal materials. These restrictions may be set by the donors of a collection who have intimate knowledge of "content and of the people represented" and wish to protect their privacy.[183] Restrictions may be set for a fixed number of years, or for the duration of an individual's lifetime. However, Sara Hodson has noted, it has become increasingly common in the past several decades for archivists to assume responsibility for protecting "the privacy rights of

[180] "Code of Ethics for Archivists," Society of American Archivists.
[181] Nissenbaum, "Privacy as Contextual Integrity."
[182] Danielson, *The Ethical Archivist*.
[183] Sara Hodson, "Private Lives: Confidentiality in Manuscript Collections." *Rare Books & Manuscripts Librarianship* 6, no. 2 (1991): 108.

individuals who are authors, addressees, or subjects of modern manuscript materials."[184] For Hodson, "the message is clear: it is in our hands to safeguard the privacy of those who cannot do so themselves."[185] And yet, she cautioned, "when the case for restriction is not clear-cut, there can be a danger that the curator's values may be imposed on the material. The curator or archivist must seek to protect individuals' privacy without engaging unwittingly in censorship."[186] Employing restrictions as a means of protecting privacy has significant implications for the access and use of archival records, a tension familiar to many archivists.

Access and Use is an area within the Code of Ethics for Archivists that determines how archivists make records available to their users. "Recognizing that use is a fundamental reason for keeping archives, archivists actively promote open and equitable access to the records in their care within the context of their institutions' missions and their intended user groups."[187] The SAA Code continues, to promote use of archives, archivists must "minimize restrictions and maximize ease of access." Archivists must thus establish a careful balance between protecting the privacy of record creators and subjects, and minimizing restrictions to ensure equitable access for all of their constituents. The International Council on Archives' Code of Ethics summarizes this balancing act more succinctly, stating that "archivists should respect both access and privacy, and act within the boundaries of relevant legislation."[188] Tension between this ethical area and that of privacy has been well documented in the archival literature.

184 Hodson, "Private Lives," 110.
185 Hodson, "Private Lives," 110.
186 Hodson, "Private Lives," 110.
187 "Code of Ethics for Archivists," Society of American Archivists.
188 " Code of Ethics," International Council on Archives, https://www.ica.org/sites/default/files/ICA_1996-09-06_code%20of%20ethics_EN.pdf.

As with so many aspects of archival work, there are no unilateral rules under which to make these decisions; instead, this work requires interpretation, research, and communication. Ethical issues and treatments will vary from collection to collection. And, as Rogers and Duranti have argued, the ethical codes provided by professional organizations cannot provide specific, practical guidelines, and in fact are often "aspiration, and therefore difficult to enforce."[189] Further, ethical concerns may potentially stand in conflict with one another, requiring archivists to make difficult, critical decisions.

3.6 Personal Records as an Archival Genre

Personal records can be "a reflection of who we are as a people - our collective memory," Robert McDonald has suggested. Indeed, "if we are to understand our history, and hence ourselves, we must find ways to preserve this patrimony."[190] Joan Schwartz and Terry Cook have written that "memory, like history, is rooted in archives… Archives contain the evidence of what went before."[191] While archives are sites of power, they can also serve to document a range of voices, experiences, and memories. The preservation of personal, as well as organizational, records is essential to supporting a micro or "bottom-up" approach to historical research. Margaret Hedstrom has argued that to equate archives and collective memory is overly simplistic, and that instead, "archivists could build a more compelling case for the social value of archives by enumerating and

189 Rogers and Duranti, "Ethics in the Cloud," 3.
190 Robert A.J. McDonald, "Acquiring and Preserving Private Records: Cultural Versus Administrative Perspectives," *Archivaria* 38 (1994): 162
191 Joan Schwartz and Terry Cook, "Archives, Records, and Power: The Making of Modern Memory," *Archival Science* 2 (2002): 18.

investigating the conditions and circumstances where archives are instrumental in forming, reviving, or transmitting a sense of shared experience."[192] Indeed, this research considers these very factors and the role of archives and archivists in transforming "evidence of me" into "evidence of us." The preceding discussion of personal records within the archival field establishes a foundation for this work.

Personal records are not limited to any particular formats or modes of creation, as McKemmish has suggested; indeed, any memory or narrative that is given a physical form might be found within the personal archive.[193] Personal records are created *by* individuals and *about* individuals, as Williams has argued.[194] Even organizational records created for official purposes might be personal for some users. If our understanding of personal records is one that allows for this degree of capaciousness – suggesting that any or all records *could* be personal – then we might ask what utility the terms "personal records" or "personal" archives have. While recognizing the potential for the personal in all records, this dissertation begins from the perspective that personal *records* are, broadly, those records that are created and maintained by individuals for reasons that are not mandated by anyone else. At the same time, records of an official nature may contain elements of the personal, or may gain personal significance over time, and find their place within aggregated personal *archives*.

In reviewing the archival literature, the lack of mandate and resistance to standardization of personal archives emerge as a defining characteristic of these records. Personal records are owed to no one. There is, in most cases, no legal or institutional mandate requiring that they be collected.

192 Margaret Hedstrom, "Archives and Collective Memory: More than a Metaphor, Less than an Analogy," in *Currents of Archival Thinking*, 176.

193 McKemmish, "Evidence of Me," 175.

194 Williams, "Personal Papers: Perceptions and Practices," 56.

Rather, they have traditionally been offered to, or solicited by, archives for their research potential. The spontaneity of their creation, their unruliness, and their resistance to standardization and easy categorization are the defining characteristics of personal records. These same characteristics are what make them so difficult to define.

4.0 Personal Records Bound to Local Digital Devices

The records discussed in this chapter have been created and stored locally on personal computers, primarily between the 1980s and the early 2000s. They have been acquired by archives and other memory institutions on physical storage media, including floppy disks, hard drives, and complete personal computers. The collections examined in the following pages include Toni Morrison's born-digital materials at Princeton University; Salman Rushdie's born-digital materials at Emory University; Susan Sontag's born-digital materials at the University of California, Los Angeles; the Deena Larsen collection at the University of Maryland; and the born-digital materials of Rafael Fajardo at the University of Colorado in Boulder. For complete descriptions of the scope and content of each collection, and specifically their born-digital components, see Appendix A.1. These collections were generated by writers, artists, and academics, and have been collected and preserved on the basis of their cultural significance and potential research value. As such, they support arguments summarized in the previous chapter, which suggest that personal records are often collected in order to serve the research interests of an institution's users and that the materials of writers and other similar cultural figures tend to be well- or over-represented in archival repositories.[195] The following exploration of these collections is organized according to this dissertation's thematic lenses: Materiality; Custody and Control; and Privacy and Publicity, and will describe the connections between these areas.

195 Powell, "The Collecting of Personal and Private Papers in Australia."

4.1 Introduction

The archival literature from the 1980s and 1990s suggests both a sense of anxiety about the challenges posed by digital records and the resolution to approach those challenges head on.[196] Even as born-digital records have become increasingly common in archives over the past forty years, many archival scholars and practitioners have observed that they continue to occupy a specialty or niche space within the profession, which continues to center paper-based records in theory, practice, and instruction. Cyndi Shein, for example, has observed that "in spite of the growing prevalence and importance of unique-born-digital resources in contemporary archives, many archival repositories have yet to responsibly address their born-digital holdings, citing lack of funding, time, and expertise as the main impediments."[197] This holds particularly true for personal digital records, which are impacted both by the historical exclusion of personal records in the profession, and by the need to adapt practices designed to suit paper-based records. Writing in the late 1990s, Adrian Cunningham observed the slow adoption and incorporation of born-digital workflows among personal papers archivists, particularly in comparison to archivists working with organizational records.[198] More recently, John Langdon has reflected on this ongoing state of affairs, suggesting that archivists responsible for government or business records have historically been more responsive to the challenges of collecting digital records in part because computers had been in use within these institutions for years, and in some instances decades, by

196 See Thomas Elton Brown, "The Society of American Archivists Confronts the Computer," *American Archivist* 47, no. 4 (Fall 1984): 366-282; Edward Weldon, "Archives and the Challenge of Change," *American* Archivist 46, no. 2 (Spring 1983): 125-134; Taskforce on Archiving Digital Information, *Preserving Digital Information*. The Commission on Preservation and Access and Research Libraries Group, 1996, https://www.clir.org/pubs/reports/pub63/.

197 Cyndi Shein, "From Accession to Access: A Born-Digital Materials Case Study," *Journal of Western Archives* 5, no. 1 (2014): 1.

198 Adrian Cunningham, "Waiting for the Ghost Train," *Archival Issues,* 24, no. 1 (1999): 55-64.

the time personal computers became available to individuals.[199] As a result, archivists working primarily with organizational records had been addressing the problem of digital records for years by the time digital records came into the custody of manuscript curators or personal papers archivists. It is helpful to consider the development of the personal computer in the 1980s and its gradual incorporation into everyday life in order to contextualize approaches to working with digital records in personal archives.[200]

As Dutton, Rogers, and Jun have explained, the personal, or home, computer "is 'personal' in that its use is intended for the individual owner (and perhaps his or her family and close friends)" rather than for primarily professional purposes.[201] Throughout the 1980s, it became increasingly common to find personal computers, often shared by the entire family, in homes throughout the United States. A statistical brief from the U.S. Bureau of the Census, entitled "The Growing Use of Computers," surveyed trends in computer use and ownership during this period, reporting that from 1984 to 1989 home computer ownership had grown from 8 percent to 15.[202] On average across all demographics included, the Census reported that 1 in 6 adults was in possession of a personal computer.[203] A special study of home computer ownership in August 2000 reported that at that time, "54 million households, or 51 percent [in the United States], had one or more computers, up from 42 percent in December 1998."[204] As the ownership and use of personal computers grew steadily, new programs and applications became available to consumers,

[199] John Langdon, "Describing the Digital: The Archival Cataloging of Born-Digital Personal Papers," *Archives and Records*, 37, no. 1 (2016): 37.
[200] William H. Dutton, Everett M. Rogers, and Suk-Ho Jun, "Diffusion and Social Impacts of Personal Computers," *Communication Research* 14, no. 2 (1987): 219.
[201] Dutton, Rogers, and Jun, "Diffusion and Social Impacts of Personal Computers," 219.
[202] "The Growing Use of Computers," U.S. Bureau of the Census (1991), https://www2.census.gov/library/publications/1991/demographics/sb91-11.pdf.
[203] "The Growing Use of Computers," U.S. Bureau of the Census.
[204] Eric C. Newburger, "Home Computers and Internet Use in the United States," U.S. Bureau of the Census (2000), https://www.census.gov/prod/2001pubs/p23-207.pdf.

expanding the role of the computer in daily life. In addition to word processing - a popular and extensively researched application - Dutton, Rogers, and Jun identified teleconferencing, electronic mail, bulletin boards, and video games as common uses of the personal computer in its early years.[205] With these developments, it became possible for individuals to create and maintain a growing number of personal records on home computers. Nonetheless, compared to business and government computing, personal or home computing was a relatively small source of records.

In addition to the rate of growth of personal computer ownership and use, some archival scholars have cited the inconsistent, idiosyncratic ways in which personal records were created as a factor contributing to the reluctance of many archivists to collect and process personal records that were created digitally. Langdon notes that "the archive's lack of control over the creation and use of digital material amplifies challenges in describing and arranging personal papers."[206] Where the consistent use of software, file formats, and record creation standards associated with institutional recordkeeping could support or guide archival workflows, personal records lacked these structures and characteristics. As early as 1994, Adrian Cunningham argued for pre-custodial intervention for personal digital records, as a way of addressing this issue. Pre-custodial interventions could ensure that personal digital records were "properly created, managed, and documented in the first instance," so as to improve the ability of archivists charged with their long-term custody and care.[207] Revisiting this recommendation in 1999, he acknowledged that it had gained little traction within the field in the intervening years, and that, in fact, "the suggestion that

[205] Dutton, Rogers, and Jun, "Diffusion and Social Impacts of Personal Computers," 220. For more on word processing, see Matthew Kirschenbaum's *Track Changes*; Brian Kunde's "A Brief History of Word Processing," https://web.stanford.edu/~bkunde/fb-press/articles/wdprhist.html; or D.A. Grier's "Coming of Age [History of Word Processing," *Computer* 39, no. 4 (2006): 5-7.
[206] Langdon, "Describing the Digital," 38.
[207] Adrian Cunningham, "The Archival Management of Personal Records in Electronic Form: Some Suggestions," *Archives & Manuscripts* 22, no. 1 (1994).

personal records archivists should seek to become actively involved in the records creation process" had been one of the major points of his earlier article to which archivists objected.[208] Cunningham cited the most common reasons given by archivists who opposed pre-custodial intervention for personal digital records, including the belief that from a practical standpoint, such interventions would be prohibitively labor-intensive.[209] Further, Shirley Spragge argued that it would not be possible "to discern early when achievements and activities make records worthy of preservation," making pre-custodial interventions impractical.[210] While Cunningham acknowledged the validity of this perspective, he argued that it would nonetheless be worthwhile to attempt to shift "the archival appraisal/selection decision closer to the time of records creation" when possible, given the fragile nature of digital records.[211] Cunningham additionally noted that some resisted pre-custodial interventions from a Jenkinsonian perspective, suggesting that the involvement of archivists early in the process of record creation would lead to "self-conscious and unnatural recordkeeping practices," thus compromising their ability to provide "objective truth."[212] To this point, Cunningham cited Terry Cook's argument that "records are contingent and need to be understood in the full context of their creation," a perspective that allows that personal records may indeed be self-conscious, but that they can still be valuable sources of evidence when understood in context.[213]

[208] Cunningham, "Waiting for the Ghost Train."
[209] Cunningham, "Waiting for the Ghost Train."
[210] Shirley Spragge, "The Abdication Crisis: Are Archivists Giving Up their Cultural Responsibility?" *Archivaria* 40 (1995): 176.
[211] Cunningham, "Waiting for the Ghost Train."
[212] Cunningham, "Waiting for the Ghost Train."
[213] Terry Cook, "Electronic Records, Paper Minds: The Revolution in Information Management and Archives in the Post-Custodial and Post-Modernist Era," *Archives & Manuscripts* 22 (1994), quoted in Cunningham, "Waiting for the Ghost Train."

While some evidence of growing support for pre-custodial intervention can be observed in the form of the personal digital archiving outreach discussed in the previous chapter, the collections discussed in this chapter suggest that this support is a more recent, and perhaps not yet pervasive, phenomenon. All of the collections discussed in the following pages were acquired years, and sometimes decades, after the time of their creation, with no pre-custodial intervention from archivists. As a result, these records, created and stored on local devices between the 1980s and early 2000s were not subject to archival processing and preservation treatments until they were acquired in the 2010s. This chapter considers both the original born-digital artifacts alongside more recent documentation from the years in which these artifacts were subject to archival intervention.

4.2 Materiality

Terry Cook has argued that in the digital environment, "the content, structure, and context of the record changes significantly from the traditional paper world. These are not stored in one physical place as on a paper page (and its stapled attachments), nor is the record itself readable by the human eye without machine and software intervention."[214] Indeed, Cook suggests, while the content of a personal digital record may resemble that of its non-digital equivalent, the context and structure are quite radically changed in the digital environment. Without the appropriate hardware and software infrastructure to support it, the digital record becomes inaccessible to users. To care for the collections discussed in this chapter, archivists draw upon a variety of collection and preservation strategies that are deeply informed by material infrastructure.

[214] Cook, "Electronic Records Paper Minds," 312-313.

The digital records within these collections exist primarily as files stored on floppy disks, compact disks, or hard drives. Toni Morrison's digital materials, for example, include files – primarily word processing documents – stored on approximately 150 floppy disks, while Rafael Fajardo's files are stored on zip disks and removable 5.25" cartridges.[215] In some instances, these records come to the archives accompanied by the hardware and software required to render them readable to the human eye. A portion of the digital records of Deena Larsen, Salman Rushdie, and Susan Sontag were deposited in the archives on external hard drives, laptops, and desktop computers.[216] Each of these five collections contains both digital files and the physical storage media on which they were received by archivists. They were all acquired by archives years, and in some cases decades, after many of their digital contents were created, and as a result, the challenges of accessing, migrating, and preserving obsolete file formats are central to the following discussion of their materiality.

These collections are examples of what Michael Forstrom has called "fugitive media," meaning digital objects for which there "has been no significant pre-custodial intervention, the digital content has not been appraised prior to acquisition, and the media is part of a collection consisting chiefly of paper-based materials."[217] With this description, Forstrom established a distinction between fugitive media within a manuscript collection and hybrid archival collections, which are understood within the field to be collections that contain both digital and non-digital

[215] "Floppy Disks," Toni Morrison Papers, Princeton University, https://findingaids.princeton.edu/collections/C1491/c4181; Rafael Fajardo collection, Media Archaeology Lab.
[216] "Collection Finding Aid," Deena Larsen Collection, Maryland Institute for Technology in the Humanities, https://archive.mith.umd.edu/larsen/about/about/index.html; "Series 11: Born Digital Materials," Salman Rushdie Papers, Emory University, https://findingaids.library.emory.edu/documents/rushdie1000/; "Series 20: Digital Materials," Susan Sontag Papers, University of California, Los Angeles, https://oac.cdlib.org/findaid/ark:/13030/kt2489n7qw/.
[217] Michael Forstrom, "Managing Electronic Records in Manuscript Collections: A Case Study from the Beinecke Rare Book and Manuscript Library," *American Archivist* 72 (Fall/Winter 2009): 461.

records in more equal quantities.[218] While the term hybrid archives may call to mind an integrated recordkeeping system that has produced both paper and digital records of archival value, fugitive media connotes digital media that have been added, perhaps haphazardly among primarily paper or other non-digital records in a collection. The Society of American Archivist's Dictionary of Archives Terminology notes that "fugitive" is also used to refer to materials "not held by the designated archives or library charged with their preservation."[219]

As the examination of this chapter's objects of study suggests, all digital records in an archival acquisition can be understood to be fugitive in some sense. This term can be used effectively to describe records acquired on digital storage media, such as a floppy disk or external hard drive, that have been separated from the digital environment or infrastructure in which it was originally used; in other words, separated from its original context of creation. The archival acquisition of files on a floppy disk or hard drive, without the hardware and software required to access the information they contain thus renders them fugitive in an additional sense. The files on Toni Morrison's floppy disks entitled "WORD PERFECT DOCS," for example, require WordPerfect, a word processing software popular in the late 1980s and early 1990s, in order to be read in their original technological context.[220] Files on the disk "WP Documents / Biblio.DBF" would require WordPerfect software as well as a program capable of reading the dBASE Table file format.[221] Structure and context are necessary in order to make content accessible. Separated from the software and hardware environments that render these files, they become fugitive. The

[218] "Appraisal Issues," The Paradigm Project Workbook on Digital Private Papers (2008), http://www.paradigm.ac.uk/workbook/appraisal/appraisal-issues.html.
[219] "Fugitive," *Dictionary of Archives Terminology*, https://dictionary.archivists.org/entry/fugitive.html.
[220] "Floppy Disks," Toni Morrison Papers, Princeton University.
[221] "dBASE Table File Format (DBF)," Sustainability of Digital Formats: Planning for Library of Congress Collections. https://www.loc.gov/preservation/digital/formats/fdd/fdd000325.shtml

digital materials explored in this chapter are fugitive in either one or both of the meanings advanced above.

Table 4 "Fugitive" properties of collections

Collection	Collecting Institution	Collection is Primarily Paper-based	Digital Records are Separated from Computing Environment
Rafael Fajardo Digital Materials	University of Colorado, Boulder	No	Yes
Deena Larsen Collection	University of Maryland	No	No
Toni Morrison Papers	Princeton University	Yes	Yes
Salman Rushdie Papers	Emory University	Yes	No
Susan Sontag Papers	University of California, Los Angeles	Yes	No

Morrison's digital records are fugitive in multiple sense of the term. Morrison's digital materials are organized into two series within her collection, constituting 134 folders within the 332-box collection, and totaling 2,051 individual files.[222] The digital materials constitute only a very small portion of the overall collection. In their account of processing Morrison's papers, archivists Elena Colon-Marrero and Allison Hughes described their experience of "discovering" the 150 floppy disks in the collection, a narrative that suggests the lack of pre-custodial

[222] Toni Morrison Papers.

intervention cited in Forstrom's criteria for fugitive media.[223] Additionally, these disks were separated from the hardware and software environments used to create and access them, as described above.

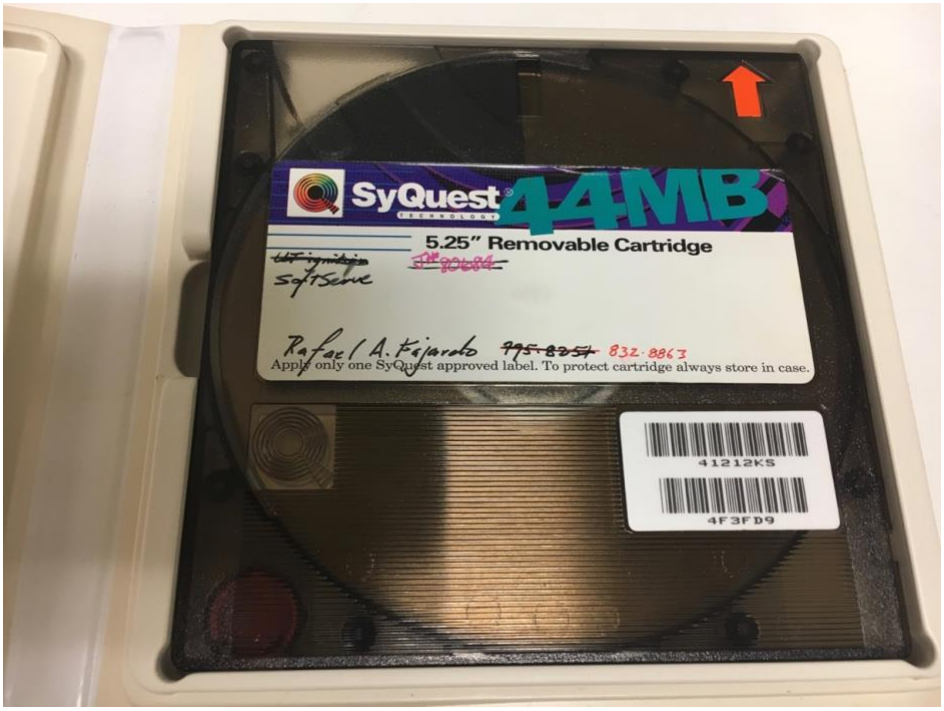

Figure 2 Rafael Fajardo's 5.25" removable disk

The collections of Deena Larsen and Rafael Fajardo contain primarily digital records, with a smaller portion of non-digital materials, a difference between these collections and Forstrom's definition of fugitive media. A majority of Fajardo's records are stored on 5.25" removable cartridges (Figure 2), and are currently inaccessible due to the absence of the hardware required to

[223] Elena Colon-Marrero and Allison Hughes, "Toni Morrison's Born Digital Material," August 26, 2015. https://blogs.princeton.edu/mudd/2015/08/toni-morrisons-born-digital-material/.

read them.[224] Larsen's collection, however, does include much of the hardware and software required to access many of her files, a characteristic shared by the digital materials of Susan Sontag and Salman Rushdie.[225]

The digital materials within the collections of both Sontag and Rushdie can be understood as fugitive in the sense that they comprise a small component of the primarily non-digital collection, and are generally confined to a single, digitally-focused series.[226] However, in both of these instances, the complete personal computers of each creator, including not only digital files but the hardware and software required to run them, were acquired with their papers.[227] As a result, the files are not separated from their original contexts of creation in the same ways that they might be if individual files had been received on removable storage media.

The acquisition of personal computers within an archival collection represents both an innovative approach to collecting digital records and a recognition of the vital role of the computing environment, which provides valuable context for digital records. When the archivists at Emory University's Manuscripts, Archives, and Rare Books Library (MARBL) began their work with Rushdie's born-digital materials, they found few comparable archival acquisitions to use as a model for their work.[228] Lacking clear precedents, a Born Digital Archives (BoDAR)

[224] "Data Storage Devices," Media Archaeology Lab, https://mediaarchaeologylab.com/catalogue/peripheral/data-storage-devices/. It is also worth acknowledging that, given the specific mission of the MAL, they are among the sites most likely to obtain obscure hardware and software required for accessing records in obsolete formats and on obsolete storage media.
[225] Matthew Kirschenbaum, "About the Deena Larsen Collection," The Deena Larsen Collection (2009), https://archive.mith.umd.edu/larsen/about/about/index.html.
[226] Susan Sontag papers, https://oac.cdlib.org/findaid/ark:/13030/kt2489n7qw/; Salman Rushdie papers, https://findingaids.library.emory.edu/documents/rushdie1000/
[227] Jeremy Schmidt and Jacquelyn Ardam, "On Excess: Sontag's Born-Digital Archive." LA Review of Books. https://lareviewofbooks.org/article/excess-susan-sontags-born-digital-archive/. Dan Rockmore, "The Digital Life of Salman Rushdie," The New Yorker, https://www.newyorker.com/tech/annals-of-technology/digital-life-salman-rushdie.
[228] Laura Carroll, Erika Farr, Peter Hornsby, and Ben Ranker, "A Comprehensive Approach to Born-Digital Archives," *Archivaria* 72 (2011): 61-92.

working group was formed and tasked with the development of a strategy for preserving and providing access to the contents of Rushdie's personal computers. Laura Carroll, Erika Farr, Peter Hornsby, and Ben Ranker, all members of the BoDAR working group, have documented their experiences processing this born-digital portion of the collection in the event that it may serve as a point of entry for other institutions that acquire complete personal computers. They have suggested that, "as the collections we receive no longer contain just one or two floppy disks, but rather may include complete operating systems and hard drives," archivists will be pushed to adapt and expand practices designed first for paper records then stretched to accommodate aggregates of individual files, or develop new practices altogether.[229] The introduction of desktop computers, laptops, mobile phones, and hand-held devices into the archives, they assert, brings with it "a transformation of accessioning procedures, processing practices, preservation tactics, and research service approaches."[230] Carroll et al.'s discussion of the Rushdie born-digital materials functions as both project documentation and a call to action: as the technologies of personal record creation change, archivists must respond by embracing new methods if they are to preserve these digital cultural heritage materials.

During the same period in which the BoDAR team was undertaking the project of processing Rushdie's born-digital materials and preserving his complete computing environment, Matthew Kirschenbaum was addressing similar issues at the University of Maryland's Institute for Technology in the Humanities (MITH). In a 2009 statement about the acquisition of Deena Larsen's personal archives, which contained a number of Larsen's personal computers in addition to papers and files on removable storage media, he acknowledged a similar shift in the contents of

[229] Carroll et al., "A Comprehensive Approach to Born-Digital Archives," 62
[230] Carroll et al., "A Comprehensive Approach to Born-Digital Archives," 62

archival collections. Kirschenbaum described Larsen's best-known work, Marble Springs, as a "transmedia artifact." It is a term that refers to the work's multiple physical manifestations and the fact that it consists of "materials both analog and digital."[231] Marble Springs is a work of interactive fiction, originally written in HyperCard.[232] Larsen used HyperCard, a pre-World Wide Web technology, because it supported one of her primary motivations in creating this text, which was to "bring the deep hidden weaving of relationships between women in a Colorado Gold Rush town to the forefront of a reader's imagination."[233] To accomplish this creative goal would require the ability to show the "links" and "connections" between entities.[234] The version of Marble Springs that resides at MITH is an excellent example of a transmedia artifact, as it consists not only of digital files - in this case, the HyperCard disk containing the interactive fiction program - but also of original material artifacts - Larsen's original Mac Classic, the computer required to run it; a crocheted cozy created by Larsen to be draped over the Mac Classic during exhibitions; and a shower curtain bearing printouts of various screens in the text, physically plotting out their connections in a diagram.[235] Understanding Marble Springs as a transmedia artifact is suggestive of the ways in which a single artifact can be hybrid, possessed of value as both a system of digital records and a physical artifact in its own right. The very ability to read these works is dependent upon the preservation of the original working environment in which and for which they were composed. Any updates to the system can threaten the integrity of the content, making access to

[231] Matthew Kirschenbaum, "About the Deena Larsen Collection."
[232] "Marble Springs," The Deena Larsen Collection, https://archive.mith.umd.edu/larsen/about/marblesprings.html.
[233] Deena Larsen, "Artist's Statement," The Deena Larsen Collection (2007), https://archive.mith.umd.edu/larsen/items/show/165/index.html.
[234] Deena Larsen, "Artist's Statement."
[235] Larsen collection inventory. Marble Springs, as a published work of interactive fiction, exists in multiple physical copies, which require a Mac Classic to run. Another copy of the text can be found at the Media Archaeology Lab, where photos included in this dissertation where taken.

original hardware and software vital to the project of preservation.[236] Larsen's collection at MITH provides an accessible, tangible illustration of these complex relationships.

In archival collections, such as those of Larsen, Sontag, and Rushdie, the personal computer embodies both artifact and collection. As an archival object, the personal computer represents not only information in the form of aggregate files but a "complete material and creative environment" in and of itself, as Kirschenbaum and others have suggested.[237] In this type of archival collection, content is preserved in the form of digital files, and context and structure are preserved in the form of interface and interaction.[238] Collecting and preserving a complete computing environment sustains elements of the physical experience of the creator: the interactions and gestures of creating, accessing, saving, and deleting. Accessing archival records by viewing the interface used by the collection's creator provides more texture to the research experience, allowing users more insight into the interactions that occur between system and user, between file and system, or between one file and another. Through the act of preserving the complete machine and providing access to an emulation of its owner's desktop interface, as in the Rushdie born-digital materials, archivists can offer researchers an opportunity to more closely observe and engage with the infrastructure supporting born-digital records, allowing for further insight into the context of their creation and use – borrowing Kirschenbaum's description of software, this environment preserves the "logical, spatial, and imaginative" elements of the records in their original environment.[239] This method of access to archival records is arguably more

[236] Angello, Aaron. "To Archive or Not to Archive: The Resistant Potential of Digital Poetry." Text Matters, 5, (2015): 13-27.
[237] Matthew Kirschenbaum, Erika Farr, Kari Kraus, Naomi Nelson, Catherine Stollar Peters, Gabriela Redwine, and Doug Reside, "Approaches to Managing and Collecting Born-Digital Literary Materials for Scholarly Use: White Paper to the NEH Office of Digital Humanities Level 1 Digital Humanities Start-Up Grant" (2009). http://drum.lib.umd.edu/bitstream/handle/1903/9787/Born-Digital%20White%20Paper.pdf?sequence=1.
[238] Langdon, "Describing the Digital," 37.
[239] Kirschenbaum, "Software: It's a Thing."

reminiscent of a historic house museum than it is a Hollinger box, giving content and context equal footing.

By collecting complete computers and providing access to both the device itself and the original computing environment by way of emulation, archivists have the opportunity to offer researchers not only the records themselves but rich contextual information about the material conditions of a record's creation and use.[240] However, this work requires significant resources, including professional training, equipment, time, and funding. Much archival literature about born-digital, locally-stored records emphasizes the fragility of these materials as a primary concern.[241] The obsolescence of hardware, software, and file formats are frequently discussed as threats to future access.[242] However, the potential for physical storage media to break or to become lost is also a concern.[243] Kirschenbaum observes that all of the digital materials in the Deena Larsen collection, without exception, are "in significant jeopardy, since software formats and physical devices are so fragile and vulnerable to obsolescence."[244] HyperCard, the program used for many of Larsen's significant works, for example, has not been updated since 1998; in 2004, Apple stopped ceased sales of the program.[245] The preservation of files, software, and hardware, many of which are designed to become obsolete in time, are necessary in order to maintain access to born-digital materials. As the following section will demonstrate, much of this work entails the transfer

[240] Carroll, et al., "A Comprehensive Approach to Born-Digital Archives, 80.
[241] Thibodeau, "Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years."
[242] Ricky Erway, "You've Got to Walk Before You Can Run: First Steps for Managing Born-Digital Content Received on Physical Media," *OCLC Research*, 2012, https://www.oclc.org/content/dam/research/publications/library/2012/2012-06.pdf; Rothenberg, "Ensuring the Longevity of Digital Information."
[243] Giles Slade's *Made to Break: Technology and Obsolescence in America* (Cambridge, MA: Harvard University Press, 2006) offers a thorough introduction to the myriad problems caused by planned obsolescence within the technology sector. The LOCKSS Program, based at Stanford University, provides some pragmatic measures taken by information professionals to mitigate loss through strategic data backups. See https://www.lockss.org/.
[244] Kirschenbaum, "About the Deena Larsen Collection."
[245] Samuel Arbesman, "The Forgotten Software that Inspired Our Modern World," *BBC News*, July 23, 2019, https://www.bbc.com/future/article/20190722-the-apple-software-that-inspired-the-internet.

of records from the storage media on which they are acquired into archival management systems designed for long-term preservation and access.

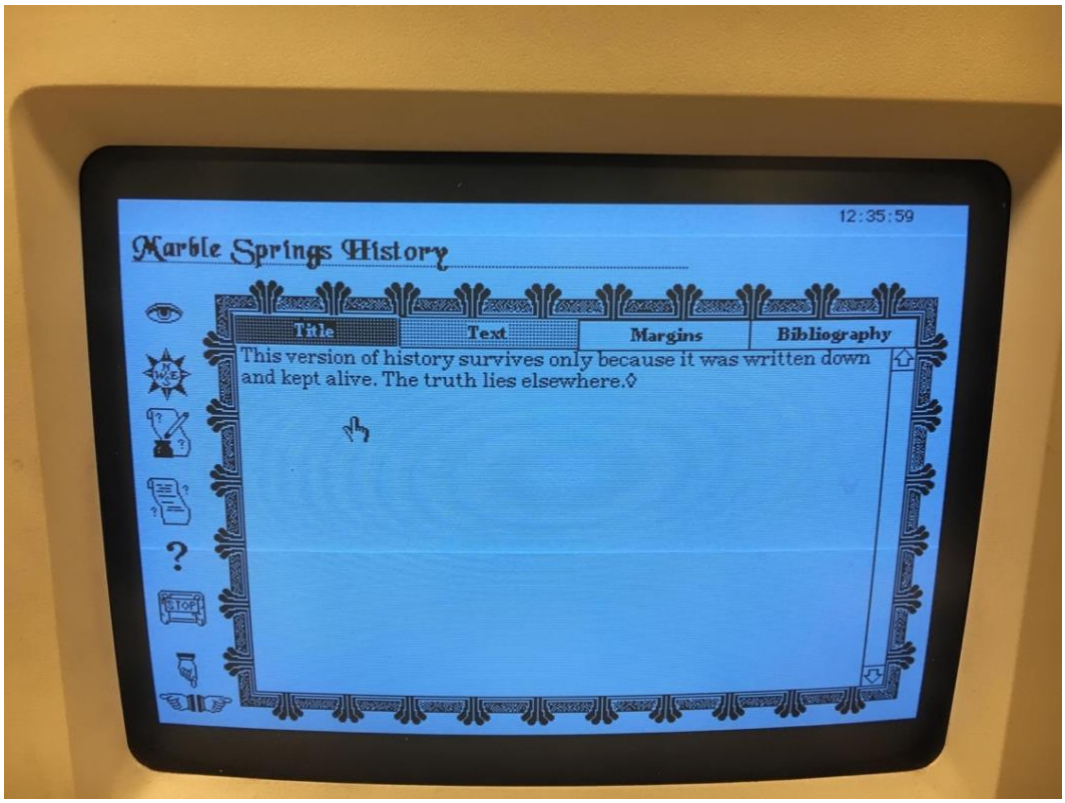

Figure 3 Marble Springs, Deena Larsen

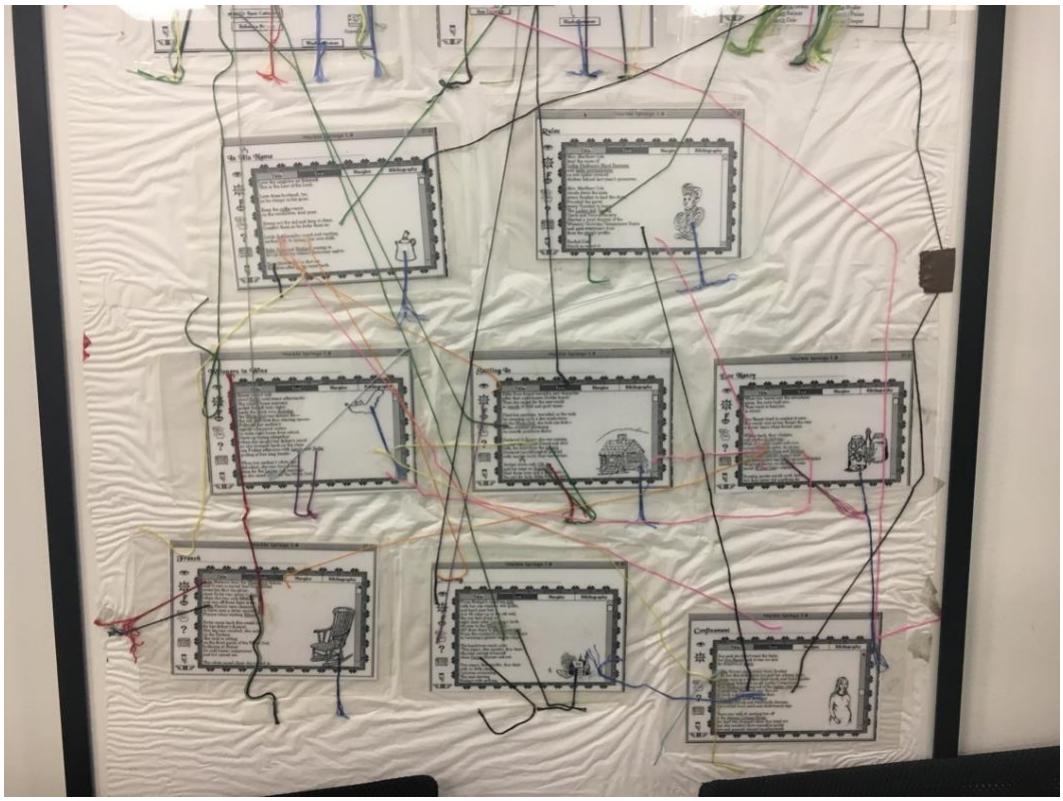

Figure 4 Deena Larsen's *Marble Springs* shower curtain

4.3 Custody and Control

Because the materials discussed within this chapter have been created and stored on local storage media and given to the archives by their creators, the proceeding discussion is specifically concerned with physical custody. As a reminder, legal custody refers to the legal ownership and responsibility for materials "regardless of their physical location."[246] Physical custody refers specifically to the "possession, care, and control" of the physical manifestation of a record.[247] The

[246] "Legal Custody," *Dictionary of Archives Terminology*, https://dictionary.archivists.org/entry/legal-custody.html.
[247] "Physical Custody," *Dictionary of Archives Terminology*, https://dictionary.archivists.org/entry/physical-custody.html.

fact that these records have been acquired on computers, hard drives, and fugitive storage media – all discrete physical objects within which a body of digital records can be created, stored, and accessed – has had a significant impact on the conditions of their custody and control. In all five of these collections, materials were deposited in the archives directly by the record creator.[248] The pages that follow will elaborate on the overall terms of legal and physical custody, as well as a discussion of the ways in which digital materials are transferred into archival systems for long-term preservation and access.

Table 5 Methods of acquiring personal digital records

Collection	Collecting Institution	Acquisition Method	Acquisition Source
Rafael Fajardo Digital Materials	University of Colorado, Boulder	Physical Transfer of Custody	Record Creator
Deena Larsen Collection	University of Maryland	Physical Transfer of Custody	Record Creator
Toni Morrison Papers	Princeton University	Physical Transfer of Custody	Record Creator
Salman Rushdie Papers	Emory University	Physical Transfer of Custody	Record Creator
Susan Sontag Papers	University of California, Los Angeles	Physical Transfer of Custody	Record Creator

While the creators of each collection discussed in this chapter formally deposited their physical materials in the archives, some took an even more hands-on approach to working with archivists. In 2014, Salman Rushdie entered into a professional relationship with Emory University

[248] In the case of the Sontag collection at UCLA, additional deposits were made after Sontag's death. These were made by the executor of her estate, following instruction by Sontag herself, and constitute a legal transfer of custody from her estate to the archives. "Provenance/Source of Acquisition," Susan Sontag papers.

with an appointment to deliver the Richard Ellman Lectures in Modern Literature, a multi-year series of biannual lectures and public readings.[249] Members of MARBL's BoDAR working group have cited this as a key factor that contributed to Rushdie's decision to deposit his papers at this institution. The author's relationship with Emory was extended through a five-year term as Distinguished Writer-in-Residence within the department, where he taught annual seminars and engaged in other university- and public-facing programs.[250] Rushdie's position and physical presence at Emory facilitated the deposit of his papers and, in particular, his born-digital records. The decision to include his complete personal computers originated with Rushdie himself, who "inquired whether or not MARBL would be interested in acquiring the computers as well as the papers" early in negotiations with MARBL archivists.[251] In their documentation of the acquisition of these materials, the archivists emphasized the importance of being able to work directly with Rushdie, a condition that allowed them to consult with him directly about any issues that arose and to "learn more about his digital life," which in turn helped them make decisions about how to process his collection.[252]

Larsen likewise wise likewise involved in, and enthusiastic about, the transfer of her materials to MITH. In an artist's statement included in the collection, Larsen wrote that she was "infinitely grateful that MITH took the time to open these before there are no computers left that will even read these files. Thank you for saving the Library of Alexandria - an entire generation of

[249] "The Richard Ellman Lectures in Modern Literature," Emory Arts, http://arts.emory.edu/about/special-programs/ellmann-lectures.html.
[250] Carroll et al., "A Comprehensive Approach to Born-Digital Archives," 63.
[251] Carroll et al., "A Comprehensive Approach to Born-Digital Archives," 63.
[252] Carroll et al., "A Comprehensive Approach to Born-Digital Archives," 65. Specifically, Rushdie was able to work with the archivists to determine the ways in which materials would be made accessible to users, and to identify sensitive or private materials for which additional restrictions to access should be applied. These issues will be discussed in greater detail in the following section, Privacy and Publicity.

works - from the flames of time."[253] However, as Kirschenbaum suggested in his remarks on the collection, due to the fragility of Larsen's computers and storage media, simply acquiring custody of these physical devices could not in itself ensure long-term access to their contents. To achieve that objective, archivists must transfer files from their original devices into archival management systems designed for long-term preservation and access. These actions constitute another means of asserting control over digital records.

While the computers and storage media included within these collections are themselves physically stored within the archives, the records that they contain are not, in most cases, accessed on those machines. Researchers accessing the records of Sontag or Rushdie do not do so at their original personal computers; nor do researchers of Morrison's archives insert her original floppy disks into a machine to access their contents. To assume custody and control of the digital files themselves, a transfer of data had to take place, often resulting in both access and preservation copies of each digital file. In four of the five collections explored in this chapter, digital records have been migrated from their original storage media into archival management systems.[254] Rushdie's materials have been made accessible to users through an emulation on a dedicated local machine in the MARBL reading room. Migration and emulation are two common means through which digital files are transferred into archival management and user access systems. Migration and emulation are methods of digital preservation, and have direct implications for the material manifestations of a digital record. However, they can also be understood as tools or methods for assuming control over those digital records.

[253] Deena Larsen, "Artist's Statement."
[254] The only collection for which migration has not taken place is that of Rafael Fajardo, at the Media Archaeology Lab, where specific equipment is needed for further action. In this instance, collection materials are physically maintained on their original storage media, located on a shelf within the Lab.

Migration is a process of moving data from one storage system or medium to another in order to ensure ongoing access to the information.[255] The transfer of digital files from Morrison's floppy disks or Sontag's personal computer into an archival management system are clear examples of this approach. In his research on digital preservation, Jeff Rothenberg identified the "physical decay of media, loss of information about the file format, encoding or compression of files, obsolescence of hardware, and unavailability of software" as central concerns.[256] Migration offers archivists "the possibility of identifying an obsolete or at-risk file format and transferring that object to a more stable, current file format."[257] However, as Andrew Pace has argued, the potential changes to which a migrated record is subject requires that archivists must "consider whether to treat digital materials as artifacts or simply as intellectual content."[258] Changes to the bitstream of a digital file have the potential to call its integrity, it's trustworthiness as a piece of evidence, into question.

Emulation is another method that has been used to provide ongoing access to some of the digital materials explored in this chapter; Rushdie's personal computer is exemplary in this area. In emulation, a current system is used to reproduce the functions of another, usually obsolete system, allowing users to access records within a version of the system in which they were created.[259] This approach is particularly useful for preserving the look, feel, and functionality of

[255] The act of copying information onto new versions of the same storage media without making any alterations is also referred to as "refreshing;" in spite of this difference in meaning, these terms are sometimes used interchangeably in archival literature and documentation. "Migration," *Dictionary of Archives Terminology*, https://dictionary.archivists.org/entry/migration.html.
[256] Jeff Rothenberg, "Ensuring the Longevity of Digital Information," *International Journal of Legal Information* 26, no. 1-3 (1998): 2.
[257] Carroll, et al., "A Comprehensive Approach to Born-Digital Archives," 77.
[258] Andrew Pace, "Coming Full Circle: Digital Preservation: Everything New Is Old Again," *Computers in Libraries* 20, no. 2 (2000).
[259] Stewart Granger, "Emulation as a Digital Preservation Strategy," *D-Lib Magazine* 6, no. 10 (2000), http://www.dlib.org/dlib/october00/granger/10granger.html.

records or systems, in addition to the information they contain. Emulation recreates the "original functionality, look, and feel by reproducing, on current computer systems, the behavior of the older system on which the document was created."[260] MARBL's archivists have employed this method, in addition to migrating individual files into an archival management system, as a means of giving users access to not only Rushdie's digital files, but the original computing environment in which they were created and used. While emulation has strong support from factions within the archival community, Rushdie's computer at MARBL is the only example of its use in the collections studied in this chapter.

Both migration and emulation entail the transfer or copying of digital records from one system to another. Though they are often discussed as digital preservation strategies, they are also tools for taking custody of records, moving them from their original storage media, along with any preservation threats inherent to those media, and into systems of archival control. In the cases explored in these chapters, migration emerges as a common means to this end. Left on their original storage media, these records and the information they contain would become inaccessible over time. Because many of these digital materials were received years after record creation occurred, however, some information was already difficult to access at the time of acquisition: for example, Morrison's WordPerfect files stored on floppy disks, Sontag's Word documents from the early 1990s, and Rushdie's faxes and Eudora email messages on an Apple Performa.

Given the challenges of accessing digital records created in obsolete file formats, with obsolete software and hardware, Kirschenbaum has suggested that a central task for archivists and other stewards of digital collections in the coming years will be the development of new tools and

[260] "Information Management: Challenges in Managing and Preserving Electronic Records: GAO-02-586," United States, Government Accounting Office (2002): 45. https://www.gao.gov/products/GAO-02-586.

best practices which can ensure that data dependent upon fragile or obsolete systems, such as the digital materials in the Larsen collection, can be safely transferred to contemporary and future systems and repositories.[261] Indeed, the collections examined in this chapter reveal a need for such tools and best practices – but they also provide evidence of the adoption of existing tools and practices from other fields to meet these needs.

Specifically, digital forensics tools and methods play a prominent role in processing the born-digital materials in several of these case studies. Digital forensics originated in the fields of law enforcement, computer science, and national defense.[262] As a field, it refers to a suite of "tools and methods for copying and analyzing all of the digital information from a physical medium in such a way that ensures the integrity and authenticity of the information are preserved."[263] A definition generated at the First Digital Forensics Research Workshop in 2001 defines the field of Digital Forensic Science as:

> "The use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation, and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations."[264]

Elements of this definition are reminiscent of aspects of archival work, and indeed, even share much of the archivist's vocabulary. As Kirschenbaum, Ovenden, and Redwine have argued,

[261] Kirschenbaum, "About the Deena Larsen Collection."
[262] Kirschenbaum, Ovenden, and Redwine, "Digital Forensics and Born-Digital Content in Cultural Heritage Collections," 1.
[263] "Digital forensics," *Dictionary of Archives Terminology*, https://dictionary.archivists.org/entry/digital-forensics.html.
[264] "A Road Map for Digital Forensic Research," *Proceedings of The Digital Forensic Research Conference*, Utica, NY, August 7-8, 2001, https://dfrws.org/sites/default/files/session-files/a_road_map_for_digital_forensic_research.pdf, 16.

> "The same forensics software that indexes a criminal suspect's hard drive allows the archivist to prepare a comprehensive manifest of the electronic files a donor has turned over for accession; the same hardware that allows the forensics investigator to create an algorithmically authenticated 'image' of a file system allows the archivist to ensure the integrity of digital content once captured from its source media; the same data-recovery procedures that allow the specialist to discover, recover, and present as trial evidence an 'erased' file may allow a scholar to reconstruct a lost or inadvertently deleted version of an electronic manuscript – and do so with enough confidence to stake reputation and career."[265]

Documentation from the Rushdie, Sontag, and Morrison collections supports this claim, and suggests that digital forensics tools factored significantly into the work of establishing intellectual and physical control over records as each collection was processed.[266]

In their account of the technical processing of Morrison's born-digital materials, Elena Colon-Marrero and Allison Hughes reported that files from the 3.5" floppy disks (which comprise approximately 80% of the disks) were transferred to the archival management system with the use of a write blocker and a Forensic Recovery of Evidence Device (FRED).[267] Later, the archivists obtained a FC5025 USB 5.25" floppy controller to gain access to the files on Morrison's 5.25" disks. The FRED, developed by Digital Intelligence, and the FC5025, developed by Device Side Data, are both tools used for the purpose of creating disk images of digital objects.[268] Disk images are high quality copies, which maintain a faithful rendering of the digital object down to the bit level. Bit-level integrity, or fixity, is a central concern in the migration of transfer of digital files from one storage device or system to another. Fixity assures a user of a file that it is indeed the

[265] Kirschenbaum, Ovenden, and Redwine, "Digital Forensics and Born-Digital Content in Cultural Heritage Collections," 2.
[266] Erika Farr and Dorothy Waugh, "Salman Rushdie Archive," BitCurator Consortium, https://www.bitcuratorconsortium.org/case-study/salman-rushdie-archive; "Save Unique, At-Risk Digital Content," OCLC Member Stories, https://www.oclc.org/en/member-stories/ucla.html; Elena Colon-Marrero and Allison Hughes, "Toni Morrison Collection," https://www.bitcuratorconsortium.org/case-studies/toni-morrison-collection.
[267] Colon-Marrero and Hughes, "Toni Morrison's Born-Digital Material."
[268] "FRED," Digital Intelligence, https://digitalintelligence.com/products/fred/; "FC5025" Device Side Data, http://www.deviceside.com/fc5025.html.

same, unchanged file and that it was not altered, corrupted, or changed in transfer.[269] It is important not only for the purpose of preserving material integrity but for preserving a trustworthy chain of custody. Digital forensics tools are immensely useful to archivists in this respect; however, their original functions in law enforcement mean that they have features and capabilities that extend beyond the preservation of data integrity.

The UCLA Library Special Collections received Sontag's digital materials on two external hard drives; because the drives contained the files transferred from her computers, rather than the original computers themselves, no disk images had to be made.[270] However, digital forensics tools and strategies were nonetheless useful to archivists. All digital files, including emails, were processed using Forensic Toolkit (FTK) 6.0.3.5. According to the collection's processing information, FTK was also used to generate reports that included the original file paths for restricted and deleted materials.

In their documentation of the processing of Rushdie's computers, MARBL archivists noted that the creation of disk images was integral to data retrieval and the subsequent emulation of Rushdie's personal computing environment.[271] With the disk images created, the BoDAR team was able to begin its comprehensive documentation of Rushdie's 12,205 MB of data – approximately 11,2350 user-generated files – in order to determine which materials would then be made accessible to researchers.[272] While digital forensics tools enable the recovery of deleted files, the BoDAR team made the decision not to recover deleted files on Rushdie's computers, largely due to the author's expressed concerns about privacy, though they noted that this decision would be

[269] "Fixity and Checksums," Digital Preservation Coalition, https://www.dpconline.org/handbook/technical-solutions-and-tools/fixity-and-checksums.
[270] "Processing Information," Susan Sontag papers.
[271] Carroll, et al., "A Comprehensive Approach to Born-Digital Archives," 71.
[272] Carroll, et al., "A Comprehensive Approach to Born-Digital Archives," 71.

made on a collection-by-collections basis in the future, in order to "benefit the donor as well as MARBL and its researchers."[273] However, Julia Kim has written, with Rushdie's support, the complete disk images of Rushdie's computers, including file paths to deleted content, search history, and cached Internet content, have been preserved in a dark archive, separate from user access copies of the data, at MARBL.[274]

While digital forensics tools and methods may be useful to archivists seeking to take control of digital records acquired on physical media, they raise ethical issues that archivists must address. Particularly when used to process the personal digital archives of private individuals, the application of methods drawn from law enforcement can be troubling. Specifically, the ability of digital forensics tools to locate and recover deleted materials – one of their primary functions in the context of law enforcement – raise issues that archivists must address if they are to rely on these tools. Gareth Knight has suggested that for archivists, this kind of recovery may be "driven by a desire to locate abandoned or previous versions of works that the creator discarded, or retrieve contextual information that provide an insight into the user's information creation processes."[275] This approach raises serious questions – for example, does recovering deleted files undermine the principle of original order? Kirschenbaum, Ovenden, and Redwine approach this issue from another angle, suggesting data recovery with digital forensics may allow archivists to recover files that were "lost or inadvertently deleted," rather than those that were intentionally destroyed. In either scenario, data recover raises urgent questions about treatments of privacy in personal

[273] Carroll, et al., "A Comprehensive Approach to Born-Digital Archives," 69.
[274] Julia Y. Kim, "Casting a Shadow, Coming out of the Dark: Digital Forensics with Personal Legacy Media," Master of Arts Thesis, Moving Image Archiving and Preservation Program, New York University, September 2014, 56.
[275] Gareth Knight, "The Forensic Curator: Digital Forensics as a Solution to Addressing the Curatorial Challenges Posed by Personal Digital Archives," *International Journal of Digital Curation* 7, no. 2 (2012): 56.

archives. These issues will be explored in the following section, in addition to other approaches to privacy in personal digital archives.

4.4 Privacy and Publicity

While many of the decisions that must be made in order to safeguard private information in personal digital archives are the same as those that must be made for non-digital collections. The transfer of personal records from the custody of the record creator to that of an archives, where it will be preserved and made accessible to new users, is a shift in context that brings with it inherent privacy violations. The deposit of digital records stored on fugitive media is no different, and yet it raises some new concerns that are specific to the digital environment in which they were created and stored. Some new concerns that are raised are specific to the acquisition of the complete personal computer, a capacious storage device, containing vast quantities of data within a system that can make it difficult for both its original user and the archivist receiving it to ascertain precisely what is stored within its various folders and directories. Beagrie has observed that steady increases in processing power and simultaneous decreases in the cost of computer storage have produced a condition in which many digital records are created, and perhaps forgotten, but nonetheless retained because there is little or no need to delete them.[276] As a result, one's personal computer may be host to a large and complex collection of files, and knowing what is included among them is difficult. Indeed, as the previous section demonstrated, with the aid of digital forensics tools, this may even include evidence of files that have been deleted.

[276] Beagrie, "Plenty of Room at the Bottom?"

Margaret Hedstrom's research on digital recordkeeping systems suggests that fears about potential privacy violations often stem from a sense that within a complex digital system discrete pieces of information can be pieced together in such a way that they would reveal much more about an individual than could any of those pieces of information individually.[277] This concern may be exacerbated by the idiosyncratic, inconsistent practices employed by many individuals in their own experiences with personal recordkeeping, particularly in the digital space. A 2011 survey of 110 writers conducted by Devin Becker and Collier Nogues suggested that their personal digital archives largely consisted of "poorly managed, highly distributed, and unsystematically labeled files, representing works of writing in myriad versions and in various stages of completion."[278] These circumstances makes it all the more challenging for donors to confidently ensure that they have identified and removed all files that they do not wish to share with archivists and future researchers. Hedstrom observed that the archivist's sense of dual responsibility to both donor privacy and researcher access is exacerbated by the introduction of born-digital materials in archival collections.[279] While each of the collections discussed within this chapter may contain records of a private or sensitive nature, the acquisition of a complete personal computer requires a deep consideration of personal privacy. As these collections demonstrate, the privacy issues themselves are not necessarily unique to the digital environment. However, many of the practical strategies that archivists have employed in order to address those concerns have been developed in order to address the challenges of the digital environment. As archivists working with Rushdie's

[277] Margaret Hedstrom, "Computers, Privacy, and Research Access to Confidential Information," *Midwestern Archivist* 6, no. 1 (1981): 5-6.
[278] Devin Becker and Collier Nogues, "Saving-Over, Over-Saving, and the Future Mess of Writers' Digital Archives: A Survey Report on the Personal Digital Archiving Practices of Emerging Writers," *The American Archivist* 75, no. 2 (2012): 483.
[279] Hedstrom, "Computers, Privacy, and Research Access to Confidential Information," 5-6.

materials at MARBL have noted, protecting specific information within specific records becomes more challenging when working with large quantities of digital data. "The existence of these restrictions," they write, "shaped much of the planning and workflow for this project."[280] Indeed, the volume of records stored on a personal computer necessitate additional, resource-intensive methods of identifying and addressing potential privacy concerns.

Recognizing the potential privacy implications of born-digital records, and specifically of complete computers, members of MARBL's BoDAR working group have detailed some of the measures taken to protect private information as "routine."[281] These included the closing of Rushdie's legal and financial until after the author's death, and closing all papers relating to his family "until the death of the specific family member, or seventy years from the date of acquisition, whichever occurs first."[282] Restrictions such as these are not uncommon in archival collections containing personal records, regardless of whether they are digital or paper-based. Additional restrictions were determined through ongoing discussions between Rushdie and MARBL archivists in order to address concerns specific to this donor and collection. For example, Rushdie conveyed his intention to publish a memoir of his life under fatwa, which would draw extensively from entries made in the journals he kept throughout this period of his life; consequently, "all journals written after 1989 are restricted," until the publication of this work.[283] This particular restriction demonstrates the value of archivists and donors working closely together to establish the conditions of access and use for collection materials. Throughout the donation negotiation process, the BoDAR team reported, Rushdie took a hands-on approach to identifying and

[280] Carroll et al., "A Comprehensive Approach to Born-Digital Archives," 68.
[281] Carroll, et al., "A Comprehensive Approach to Born-Digital Archives," 68.
[282] Carroll et al., "A Comprehensive Approach to Born-Digital Archives," 68.
[283] Carroll et al., "A Comprehensive Approach to Born-Digital Archives," 68.

restricting "certain portions of his papers, as the collection included a significant amount of personal, financial, and other sensitive information."[284]

Similar traditional approaches to restricting private information can be observed in the Sontag collections. Two boxes containing the writer's private journals, for example, have been restricted until 25 years after her death.[285] An interview with Victoria Steele of UCLA's special collections from the time of the initial acquisition of the collection indicates that these privacy restrictions were conditions set by Sontag herself.[286] However, some of the approaches taken to making Sontag's born-digital materials accessible are indicative of the heightened degree of access to personal information that is possible in the digital environment.

Sontag's born-digital materials include 17,198 emails sent by the writer during her lifetime, a large corpus of correspondence.[287] Correspondence, including electronic correspondence is generally written to be read only by the sender and their recipients. Sontag's personal messages "reveal the minutiae of her friendships, her appointments, and the small forgotten details of everyday life."[288] Sontag's biographer Benjamin Moser has reflected on the experience of reading through these messages, writing that the electronic correspondence provoking a feeling of intimacy that he had not experienced previously in the course of conducting research: "reading papers and manuscripts is one thing. Looking through someone's email is quite another."[289] Moser posited the historical proximity and the often mundane, everyday nature of email as the cause of this

[284] Carroll et al., "A Comprehensive Approach to Born-Digital Archives," 68.
[285] Biber and Luker, "Evidence and the Archive."
[286] Mimi Avins, "UCLA Buys Sontag's Archive," *Los Angeles Times*, January 26, 2002, https://www.latimes.com/archives/la-xpm-2002-jan-26-mn-24914-story.html.
[287] "Born-Digital Materials," Susan Sontag papers.
[288] Katherine Biber and Trish Luker, "Evidence and the Archive: Ethics, Aesthetics, and Emotion," *Australian Feminist Law Journal* 40, no. 1 (2019).
[289] Benjamin Moser, "In the Sontag Archives," *The New Yorker*, January 30, 2014, http://www.newyorker.com/books/page-turner/in-the-sontag-archives.

uneasiness. However, the infrastructure of a digital system allows for additional forms of access that are just as likely to produce a feeling of closeness, or of uncanny access. The ability to keyword search and sift through all email correspondence for a specific word provides ready access to specific content, without the requisite, slow process of perusing each document. Jeremy Schmidt and Jacquelyn Ardam demonstrate this with a keyword search for "Annie," instantly retrieving all messages referencing or addressed to Sontag's partner Annie Leibovitz.[290] Additionally, UCLA supports the use of Muse ("Memories Using Email"), a program offered by Stanford University's Computer Science department, to analyze trends throughout Sontag's email correspondence.[291] Muse mines data in a corpus to "uncover long-forgotten topics and people across tens of thousands of messages," and uses sentiment analysis to provide users with insights into the "highs and lows" of their lives, as "captured" in their email accounts.[292] These tools enable fast access to specific details and insights not so readily accessible in non-digital collections.

The use of digital forensics tools in both the Sontag and Rushdie collections has resulted in further questions about how to approach the treatment of private information. Specifically, digital forensics tools were used to screen for private information within the digital materials in each collection. Documentation of the Sontag collection details the screening process, explaining that the records were scanned for

> "Personally Identifiable Information (PII) and other sensitive information using FTK 6.0.3.5. Files containing PII and/or other sensitive information have been permanently restricted in accordance with federal regulations and/or UCLA Library Special Collections privacy policies."[293]

[290] Schmidt and Ardam, "On Excess: Susan Sontag's Born Digital Archive."
[291] Schmidt and Ardam, "On Excess."
[292] "Muse," Stanford University, https://mobisocial.stanford.edu/muse/. Muse does state that users can run the program on their own computers "without giving up the data anywhere else," ensuring the privacy of user data. IRE-2012, Tip Sheet for Journalists, https://mobisocial.stanford.edu/muse/tipsheet.html.
[293] "Digital Materials – Processing Information," Susan Sontag papers.

This documentation describes the host of standards and policies that determine what is understood as private or sensitive information within these personal digital materials. Tools like FTK allow users to search for PII using keywords regular expressions. For example, Social Security numbers can be identified using "<\d\d\d[-]?\d\d[\-]?\d\d\d\d\> where \d represents a 'digit.'"[294] In order to be effective, this requires archivists to know exactly what kinds of information they are searching the corpus for, and how that information is structured. In this context, PII is understood narrowly, and documentation generally refers to information including Social Security numbers, financial information, and specific keywords.[295] While useful in protecting those very specific pieces of information, digital forensics tools cannot account for the emotional content of a record, or the specific contexts in which otherwise seemingly innocuous information may be sensitive or private.

Indeed, at MARBL, archivists found that digital forensics tools were insufficient for ensuring that no private or sensitive information related to Rushdie, his family, or friends would be accessible to researchers using the collection. They have explained that a condition of the acquisition of Rushdie's digital materials was that "correspondence from a select number of individuals could be opened [to researchers] only if phone numbers, fax numbers, and home addresses were redacted from the records."[296] As MARBL staff began processing the digital materials, they found that, even when focusing on structured strings of information, such as phone and fax numbers, the "time, resources, and development needed to effectively redact sensitive

[294] Yin Pan, Bill Stackpole, and Luther Troell, "Computer Forensics Technologies for Personally Identifiable Information Detection and Audits," *ISACA* 2 (2010): 5.
[295] Pan, Stackpole, and Troell, "Computer Forensics Technologies for Personally Identifiable Information Detection and Audits," 2.
[296] Carroll et al., "A Comprehensive Approach to Born-Digital Archives," 68.

information from the correspondence proved too great for the work schedule and resources" available during the first phase of the project.[297] After consulting with Rushdie, it was decided that, with the exception of selected correspondence files from his first computer, all digital correspondence would remain closed to researchers. Remaining correspondence would be made accessible to researchers only when the time and resources required to properly review and redact private information. Likewise, sections of digital materials in Sontag's archive that contain "private family information are restricted until December 2044. Other files that contain sensitive medical information, personally identifiable information, and software program files have been restricted in perpetuity pending curatorial review."[298] These examples provide some insight into the complexity of using digital forensics tools in service of protecting private or sensitive information. Even for institutions with significant resources and expertise such as Emory University and UCLA, the extensive labor required to process born-digital collections like those of Salman Rushdie and Susan Sontag with both technical precision and ethical treatment is prohibitive.

The use of digital forensics tools also requires that collection stakeholders - and specifically, donors and archivists - develop a policy for deleted records. The BoDAR working group explains their decision not to recover deleted files from Rushdie's computers, writing that because of "the nature of this collection and sensitivity of some of this material, coupled with concerns that Rushdie expressed about his privacy and the privacy of his family and friends," they determined that "data recovery would not be appropriate" for the collection.[299]

[297] Carroll et al., "A Comprehensive Approach to Born-Digital Archives," 68.
[298] "Restrictions on Access," Susan Sontag Papers.
[299] Carroll et al., "A Comprehensive Approach to Born-Digital Archives," 69.

The application of tools and methods from the field of digital forensics leads to some important questions about how privacy and private information are understood and protected. Archivists will have to make decisions, as they have with the collections discussed in this chapter, about the degree to which they will utilize these tools and defer to understandings of privacy drawn from the field of digital forensic science. Christopher A. Lee has argued that the "incorporation of digital forensics methods will also be essential to the sustainability of archives as stewards of personally identifying information," suggesting that "the same tools that are used to expose sensitive information can be used to identify, flag, and redact or restrict access to it."[300] This is an optimistic perspective. Digital forensics tools are designed to discover and to expose private information. To wield technologies designed for exposure as tools for protecting private information requires a deliberate, critical approach to working with personal archives, in which archivists consciously balance technological capabilities with ethical and compassionate approaches, balancing researcher access with the rights and wishes of record creators.

4.5 Conclusion

The collections discussed in this chapter can all be understood to be hybrid archives, consisting of both digital and non-digital records. More specifically, these records come into the archives as fugitive media. Drawing upon Forstrom's definition, these are digital media within a primarily non-digital collection, for which there has been little or no pre-custodial intervention and

[300] Christopher A. Lee, "Archival Application of Digital Forensics Methods for Authenticity, Description, and Access Provision," *Comma* 2, no. 4 (2012): 137.

no appraisal prior to archival acquisition.[301] Many of the records discussed in these pages are also fugitive in another sense: acquired as collections of files on storage media, removed from the original hardware and software environments in which they are created, they are separated from their contexts of creation. In the cases of Larsen, Rushdie, and Sontag, complete computers have been acquired by archives, maintaining digital files within the context of their original hardware and software environments. Of Larsen's collection of computers, Kirschenbaum suggested that "such hybrid, transmedia works are not anomalous but in fact typical of the kind of cultural heritage libraries and repositories will have to learn to curate and archive in the years to come."[302] Indeed, examples of digital, fugitive media within archival collections can be found in a steadily growing number of archives in addition to those discussed within this dissertation.

However, as Shein has observed, the archival literature tends to draw primarily on case studies of high-profile collections within institutions with technological resources and expertise, which may not be able to serve as useful models for smaller institutions without a robust digital infrastructure.[303] Shein has noted additionally that because of the high profiles of the creators of these collections and the capabilities of the institutions that collect them, they are often subject to highly technical and labor-intensive processing strategies, including "the emulation of the creators' computing environments and/or granular (often file-level) description of the content."[304] Emory University Library and the Maryland Institute for Technology in the Humanities at the University of Maryland are examples of such institutions, and while their digital preservation projects are "groundbreaking," they "come from a very similar and limited perspective - that of large

[301] Forstrom, "Managing Electronic Records in Manuscript Collections," 461.
[302] Kirschenbaum, "About the Deena Larsen Collection."
[303] Shein, "From Accession to Access," 3.
[304] Shein, "From Accession to Access," 3.

institutions with solid funding and excellent technical support."[305] The Rushdie digital materials at Emory University stand as a prominent case study for such an approach. The Rushdie born-digital collection represents an innovative approach to both processing and presenting a complete personal computing system. It also demonstrates many of the challenges, both technical and ethical, faced by archivists working with born-digital materials of this nature.

As the collections considered within this chapter suggest, however, even within well-resourced institutions, born-digital processing is resource-intensive. Indeed, in a 2017 presentation at the Personal Digital Archiving conference at Stanford University, Dorothy Waugh and Elizabeth Russey Roke of MARBL reflected on the project's status, acknowledging that even with their resources "moving from acquiring Rushdie's computers to providing reading room access to his files and computing environment involved a steep learning curve as archivists and software engineers grappled with how the nature of born-digital files affected arrangement, description, and access at a time when relatively few institutions were working with born-digital material."[306] They reported that in the six years since the project began, Rushdie's remaining computers have been processed, but that scaling from a digital archives project to a wider-reaching digital archives program has proven challenging.

In addition to serving as a storage repository for aggregated personal records, applications, and activity logs, the computer is itself a physical artifact to be maintained. As a physical object, the computer provides context for the records created with its assistance, while simultaneously

[305] Shein, "From Accession to Access," 2-3.
[306] Dorothy Waugh and Elizabeth Russey Roke, "Second Generation Digital Archives: What We Learned from the Salman Rushdie Project," Personal Digital Archiving conference (2017), https://personaldigitalarchiving2017.sched.com/event/9JaJ/session-2-preserving-serving-pda-at-memory-institutions

serving as an additional form of evidence in and of itself.[307] Kirschenbaum has compared the preservation of a writer's personal computer to that of a non-digital writing tool or environment, such as Jane Austen's writing table or Paul Auster's typewriters and red notebooks.[308] In her reporting on the preservation of Salman Rushdie's complete personal computer and the emulation of his computing environment, Patricia Cohen has drawn similar parallels, suggesting that for the BoDAR working group, " simulating the author's electronic universe is equivalent to making a reproduction of the desk, chair, fountain pen and paper that, say Charles Dickens used, and then allowing visitors to sit and scribble notes on a copy of an early version of 'Bleak House.'"[309] In these analogies, the personal computer is material, spatial, and evocative of the person who used it; it retains something of the aura of its original user. And yet these comparisons fall short of describing the full complexity of the object and the operations it is used to perform. While it is undoubtedly a physical object that can be taken in and appreciated visually, it is the interior world of the computer that truly activates it as an object. Ciaran Trace has drawn on the work of physicist Richard Feynmen and computer engineer Jon Stokes to offer another analogy that comes closer to accomplishing this goal, likening the personal computer to a "self-contained office part composed of numerous buildings."[310] Trace's analogy neatly illustrates the complex system of storage, retrieval, and other activity carried out within the computer, while acknowledging that the tasks that have been automated and operationalized within the personal computer are based upon actions performed by humans. "Not only has this technology annexed human work and work processes,"

[307] Edward Rothstein, "Displaying the Dinosaurs of the Digital Age," *New York Times*, September 28, 2012, https://www.nytimes.com/2012/09/29/arts/design/computer-history-museum-in-mountain-view-calif.html.
[308] Kirschenbaum, *Track Changes: A Literary History of Word Processing*, (Cambridge, MA: The Belknap Press of Harvard University, 2016), 8.
[309] Cohen, "Fending Off Digital Decay, Bit by Bit." https://www.nytimes.com/2010/03/16/books/16archive.html
[310] Trace, "Beyond the Magic to the Mechanism," 10.

she explained, "the computer has also integrated and re-imagined the attendant recordkeeping paraphernalia of records, files, file folders, file cabinets, and file rooms - all this contained within an object that has an increasingly small footprint."[311] Like an office, or like a house, the computer contains a variety of locations, functions, and objects. As both object and collection, myriad decisions about how each of these will be preserved must be made.

The specific material characteristics of a digital object determine the means through which custody and control are established. Because the collections discussed in this chapter were stored locally, the discussion of custody was closely linked to the material, physical traits of digital records. Control over the records was established through the transfer of the physical media and devices on which they were stored. It was further enacted through the transfer of digital files from the original storage media on which they were acquired and into archival management systems for long-term preservation and access. The digital materials discussed in this chapter are collections of digital files stored on obsolete media, created with obsolete software. To transfer them from their original storage media and into archival systems for preservation and access, digital forensics tools and strategies were used in most cases. Digital forensics tools have proven incredibly valuable for archivists charged with retrieving files in obsolete formats from storage devices that, in many cases, have not benefited from any pre-custodial interventions. While these tools and approaches can "advance the archival goals of maintaining authenticity, describing born-digital records, and providing responsible access," in addition to uncovering and protecting the provenance of digital records, most digital forensics tools were "not designed with archival objectives in mind."[312]

[311] Trace, "Beyond the Magic to the Mechanism," 11.
[312] Lee, "Archival Application of Digital Forensics Methods," 137.

Because the digital materials discussed in this chapter were created and stored locally on computers, and formally deposited in the archives by their creators, decisions about privacy and access restrictions were more often addressed through direct communication and negotiation with record creators and donors. Members of MARBL's BoDAR working group affirmed that "because of the great potential for content creators to shape their own digital archives, archivists and curators must work with donors before and during the acquisition process."[313] In the collections of Morrison, Sontag, and Rushdie, collection documentation suggests that input from record creators was instrumental in establishing privacy and access restrictions. While digital forensics may be useful for a range of practical tasks, this research suggests that these tools and methods are insufficient for addressing privacy concerns in an archival context.

While these collections were locally stored and acquired on physical storage media, the potential impact of the Internet was occasionally hinted at. In Rushdie's collection, for instance, Internet search histories and cached browsing history were included in his personal computers. The collections of both Rushdie and Sontag include email correspondence, further gesturing to the kinds of digital records that can be created with networked or web-based computers. Other collections, like Larsen's, are positioned at the precipice just before the popularization of the World Wide Web. In a statement about the collection, Larsen notes that, "Marble Springs debuted in 1993, the year HyperCard died," positioning her work at a moment of profound technological transition.[314] At the 1993 Hypertext Conference, she witnessed "the world-wide-web [sic] in its infancy (and of course, completely failed to recognize its significance), and bewailed the announcement that HyperCard would no longer be a supported Apple software."[315] Larsen

[313] Carroll, et al., "A Comprehensive Approach to Born-Digital Archives," 90.
[314] Deena Larsen, "Artist Statement."
[315] Deena Larsen, "Artist Statement."

characterizes HyperCard as the "Lewis and Clark expedition into the realms of what would become the Web, the first foray into the wilds before the Web became the fast and loose Wild West of the dying twentieth century."[316]

Echoing Kirschenbaum's assertion that archival work in the future will entail more and more work with transmedia, digital objects, the BoDAR working group suggests that the introduction of desktop computers, laptops, mobile phones, and other hand-held devices in archives will prompt "a transformation of accessioning procedures, processing practices, preservation tactics, and research service approaches."[317] In *The New Yorker,* Dan Rockmore surmised that Rushdie's next deposit in the archives might "include tweets and Instagram photos, which will be added to decades of his digital output."[318] The Rushdie born-digital materials, as well as the born-digital materials in other collections discussed in this chapter, certainly support this statement. The introduction of web-based personal digital records, to be discussed in the next chapter, will have a similarly transformative effect, introducing a variety of new concerns for archivists and other stewards of personal digital archives.

316 Deena Larsen, "Artist Statement."
317 Carroll et al., "A Comprehensive Approach to Born-Digital Archives," 62.
318 Rockmore, "The Digital Life of Salman Rushdie."

5.0 Personal Records on the World Wide Web

The personal records discussed throughout this chapter have been created and stored on the open World Wide Web in the 1990s and 2000s. They include static web pages and online diaries and blogs. They have since been acquired by archival repositories and other memory institutions, as well as by activist collectors working outside of traditional collecting institutions. The collections examined in the following pages include the Katie Lee Papers at Northern Arizona State University; the Mormon Missionary Collection at Brigham Young University; the Zine Web Archive at the Library of Congress; and two simultaneous efforts to preserve websites created with the GeoCities website creation and hosting service: one orchestrated by the Internet Archive, and other by the group of self-described "rogue archivists" known as Archive Team. For a more thorough introduction to each of these collections, and specifically their web-based, digital components, see Appendix A.2. Where the collections studied in the previous chapter were acquired on physical media, directly from their creators or their estates, the collections discussed in the following pages have been collected through a more varied range of methods, and from a variety of sources. In some instances, they are acquired through a formal transfer of custody from creator to archives; more frequently, however, they are acquired directly from the open web with the assistance of web archiving tools. In these respects, they represent a departure from the collecting strategies employed with locally-stored born-digital personal records. The proceeding examination of these records is organized according to this dissertation's thematic lenses: Materiality; Custody and Control; and Publicity and Privacy, and will explore the connections between these areas.

5.1 Introduction

"In terms of human culture, the Internet is still very much an infant, but someone forgot to tell this to the millions of people who have added their contributions to it," Carol Casey lamented in 1998.[319] The World Wide Web had been invented and made available to the public less than a decade earlier, but it had been quickly established as a place where information was shared, records were created, and communication took place.[320] By the mid-1990s, increasing access to this new information environment had resulted in a growing body of records – which many archivists recognized early on as new, valuable contributions to digital cultural heritage – created in new formats, stored and accessed within new infrastructure, and subject to new preservation threats.

From the time of its launch in 1990, the use of the World Wide Web accelerated swiftly, in a trajectory of growth not unlike the pattern of adoption of the personal computer documented in Chapter Four. A survey conducted by the United States Census reported that by the year 2000, 51 percent of households were in possession of a personal computer, while 41.5% of households had Internet access.[321] By 2009, those numbers had risen considerably, with a reported 76.7% of American households in possession of personal computers, with 71.1% having Internet access at home.[322] As Susanne Belovari has recently observed, throughout the 1990s and 2000s, the World Wide Web "became the location where we carry out more and more of our activities and increasingly generate primary and secondary records."[323] Archivists and information professionals

[319] Carol Casey, "The Cyberarchive: A Look at the Storage and Preservation of Web Sites," *College & Research Libraries* 59, no. 4 (1998): 309.
[320] "A Short History of the Web," CERN, https://home.cern/science/computing/birth-web/short-history-web.
[321] Eric C. Newburger, "Home Computers and Internet Use in the United States: August 2000," U.S. Bureau of the Census, https://www.census.gov/prod/2001pubs/p23-207.pdf.
[322] Thom File, "Computer and Internet Use in the United States," U.S. Bureau of the Census, https://www.census.gov/prod/2013pubs/p20-569.pdf.
[323] Susanne Belovari, "Historians and Web Archives," *Archivaria* 83 (Spring 2017): 60.

found themselves, once again, responsible for collecting and preserving the records that were being created and disseminated within a new technological infrastructure, even as that infrastructure was still in the early stages of its development.

The early web of the 1990s is often characterized as something of a wilderness - "communal, libertarian, collaborative, occasionally raucous, anti-establishment and rich in debate and discussion," as John Naughton has depicted it in his research on the evolution of the Internet.[324] Users of the web could create personal websites that documented their lives, hobbies, or special interests and share these records of their lives with others on the open web.[325] Because any user of the web could create and disseminate records in addition to accessing records created by others, many have characterized the World Wide Web as a key player in the "democratization of information."[326] Tools and services that supported the creation of websites without requisite coding skills, including GeoCities and later blogging services like Blogger and WordPress, further enabled personal record creation on the web. This infrastructure supported the creation and dissemination of records from anyone with access to a computer with Internet access, and in part, including many whose voices had traditionally been underrepresented in publishing, broadcasting, and cultural heritage. The potential historical significance of these web-based personal records was not lost on archivists, many of whom have been engaged in the preservation of these records from the early days of the World Wide Web.[327]

[324] John Naughton, "The Evolution of the Internet: From Military Experiment to General Purpose Technology," *Journal of Cyber Policy* 1, no. 1 (2016): 16.
[325] The term "open web" refers to the public-facing side of the World Wide Web, encompassing the parts of the web that are public and viewable by anyone with an Internet connection and web browser. This is sometimes also referred to as the "surface" or "crawlable" web, referring to the ability of spiders and crawlers to scrape and index its contents. See PC Magazine Encyclopedia, "Surface Web," https://www.pcmag.com/encyclopedia/term/surface-web.
[326] See Jennifer Granick, "Saving Democracy with Web 2.0," *Wired,* https://www.wired.com/2006/10/saving-democracy-with-web-2-0/; Justin Reich, "Reworking the Web, Reworking the World: How Web 2.0 is Changing our Society," *Beyond Current Horizons*.
[327] Carol Casey, "The Cyberarchive," 309.

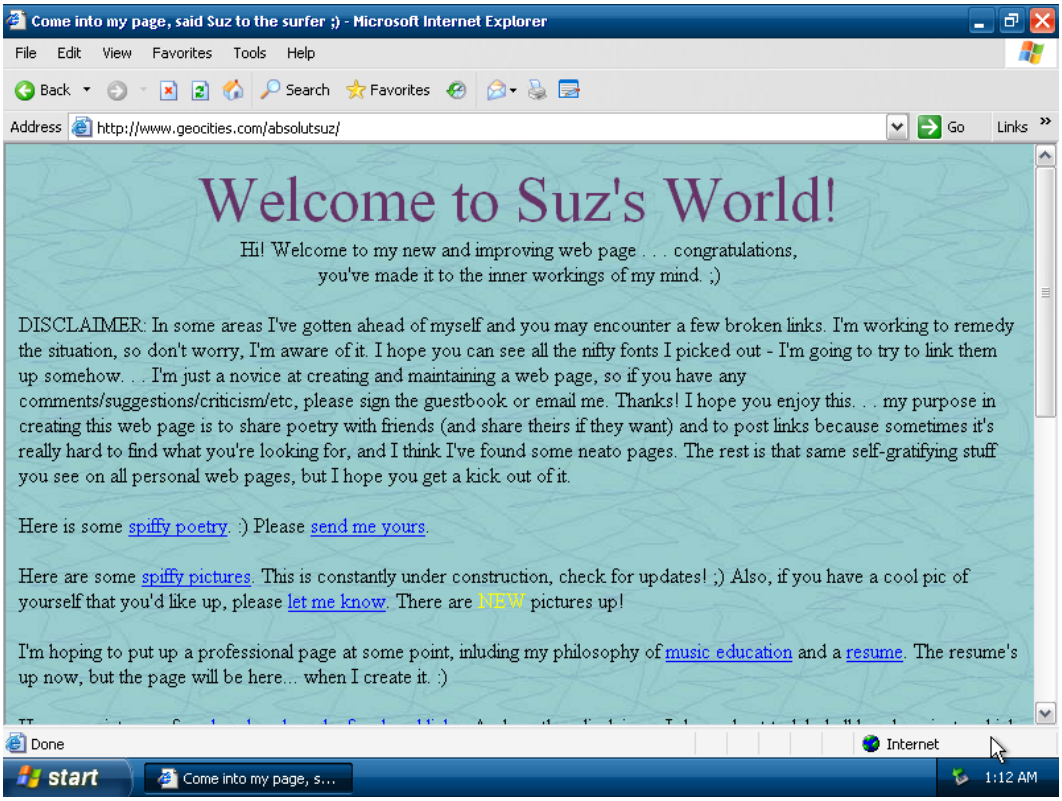

Figure 5 Screenshot "Welcome to Suz's World"

Reviewing the archival literature from the 1990s and early 2000s reveals an ongoing process of making sense of what web-based personal digital records are, and why and how to preserve them. In some instances, these records were conceptualized as extensions or descendants of non-digital record formats that have long been recognized as valuable forms of historical evidence.[328] Susan Lukesh took this approach to studying email and online correspondence, likening them to handwritten or typed letters.[329] Catherine O'Sullivan likewise compared online

328 Richard Cox, *Personal Archives and the New Archival Calling*.
329 Susan Lukesh, "E-mail and Potential Loss to Future Archives and Scholarship: The Dog That Didn't Bark," First Monday 4, no. 9 (1999).

diaries and blogs to handwritten diaries and journals.[330] These comparisons are perhaps most effective when they are used for the purpose of advocating for the evidentiary value of digital records; however, they prove somewhat less apt when drawing comparisons between the approaches use with non-digital, locally stored records and those needed to collect, preserve, and provide access to web-based personal digital records. In order to preserve the records of web-based digital cultural heritage, archivists have had to develop practical strategies and technologies that are suited specifically to the unique material, legal, and social characteristics of web-based personal records. Web archiving tools and methods have been developed, both within the archival profession and beyond it, to support the collection and preservation of web-based records.

The term "web archiving" refers to "the process of gathering up data that has been recorded on the World Wide Web, storing it, ensuring the data is preserved in an archive, and making the collected data available" for future use.[331] Like archival work more broadly, web archiving encompasses a complex system of professional activities, including appraisal, acquisition, arrangement and description, preservation, and the provision of access. These activities may be deployed in a variety of ways depending upon the specific content, context, and structure of the specific web-based records in question. The Internet Archive, which began its work in this area in 1996 with the mission of "archiving the whole Internet," is widely recognized as a progenitor of the web archiving tools and methods used throughout the field today.[332] In the years since it began its efforts to preserve the Internet in earnest, it has grown and expanded its collecting practices, and has arguably risen to the level of infrastructure within the field. In addition to its Wayback

330 Catherine O'Sullivan, "Diaries, On-Line Diaries, and the Future Loss to Archives: Or, Blogs and the Blogging Bloggers Who Blog Them," The American Archivist 68, no. 1 (2005).
331 Jinfang Nu, "An Overview of Web Archiving," *D-Lib Magazine* 18, no. 3-4 (2012), http://www.dlib.org/dlib/march12/niu/03niu1.html.
332 "About the Internet Archive," Internet Archive, https://archive.org/about/.

Machine web archiving service, which is home to one of the GeoCities collections discussed in this chapter, the Internet Archive provides the subscription web archiving service Archive-It, which has been used to build the Mormon Missionary Collection. The Library of Congress's web archiving initiatives, including the Zine Web Archive, rely upon the Heritrix web crawler, also developed by the Internet Archive.[333] Indeed, even the preservation standard file format for web archives, the WebARChive (WARC) file format, originated with the Internet Archive's ARC_IA format, which specified "a method for combining multiple digital resources into an aggregate archival file together with related information."[334] As this chapter's research will suggest, the role of the Internet Archive in the web archiving landscape cannot be understated.

Many web archiving strategies have been developed in response to a pervasive understanding of the World Wide Web as a volatile space, and of websites as inherently unstable, even ephemeral, records. As Rick Barry has observed, "web sites make records, but they do not keep records in ways that match up to sound recordkeeping requirements."[335] Both archival and mainstream bodies of literature on web-based records reveal a strong sense of fear that websites can disappear just as suddenly as they have been created, taking with them any information or evidence they might provide. Terry Kuny famously warned of the potential for a "digital Dark Ages," as the result of the rapid loss of born-digital information, with web-based digital information recognized as particularly prone to loss.[336] In a 1997 *Scientific American* special report, Internet Archive founder Brewster Kahle claimed that the average lifespan of a website

[333] "For Site Owners," Web Archiving, Library of Congress, https://www.loc.gov/programs/web-archiving/for-site-owners/.
[334] "ARC_IA, Internet Archive ARC File Format," Sustainability of Digital Formats: Planning for Library of Congress Collections," https://www.loc.gov/preservation/digital/formats/fdd/fdd000235.shtml.
[335] Rick Barry, "Web Sites as Recordkeeping and 'Recordmaking' Systems," *Information Management Journal* 38 (November/December 2004): 27.
[336] Terry Kuny, "A Digital Dark Ages? Challenges in the Preservation of Electronic Information," *International Preservation News* 17 (May 1998): 10, https://www.ifla.org/files/assets/pac/ipn/17-98.pdf.

URL was only 44 days appears to be a touchstone for this point.[337] Other studies have suggested that actual average lengths are closer to 75 or even 100 days; nonetheless, it has been widely accepted that websites are short-lived, in spite of the fact that, as Nicholas Taylor has suggested, these metrics "take for granted that we know what it means that a webpage has 'died.'"[338] Websites may be lost for a variety of reasons, including the discontinuation of hosting services, deletion by the website creator, technological error, or changes to the original URL.

The work of web archiving is the work of capturing unstable records and preserving them as stable, reliable records that can be made accessible in the long term. As several of the collections explored throughout this chapter will demonstrate, concerns about the volatility and ephemerality of web-based records, coupled with the availability of tools that support the collection of records directly from the open web, have led to somewhat radical shifts in the approaches taken to collecting and preserving web-based personal records.

5.2 Materiality

From its inception, the World Wide Web was frequently depicted as placeless, a space beyond the bounds of spatially situated, material infrastructure.[339] Websites could be accessed from any computer with an Internet connection; they seemed to be simultaneously everywhere and nowhere. Much of the physical infrastructure that supported access to the web – beyond one's own

[337] Brewster Kahle, "Preserving the Internet," *Scientific American Special Report*, http://web.archive.org/web/19970504212157/http://www.sciam.com/0397issue/0397kahle.html.
[338] Nicholas Taylor, "The Average Lifespan of a Webpage," *The Signal*, November 8, 2011, https://blogs.loc.gov/thesignal/2011/11/the-average-lifespan-of-a-webpage/.
[339] John Perry Barlow, "A Declaration of the Independence of Cyberspace," Electronic Frontier Foundation, February 8, 1996, https://www.eff.org/cyberspace-independence.

personal computer screen – was concealed from users. Though so much of this infrastructure is invisible or inaccessible, engaging with it is critical to the process of preserving websites and other web-based personal records for long-term access.[340] Depending upon the approach being taken to collecting these records, varying levels of access to their underlying infrastructure are required. Of the five collections discussed throughout this chapter, only one has been acquired by archivists as a complete, live website; the other four have been acquired as archived "snapshots" through the application of web crawling tools.

Live websites exist, much like the personal digital records discussed in the previous chapter, as files on physical storage media; however, web-based records can be accessed remotely from devices other than those on which the website files are stored. Anne Kenney and Nancy McGovern have described a website a collection of aggregated web pages that "resides on a server within an administrative context, all of which may be affected by the external technical, economic, legal, organizational, and cultural environment," much of which is unseen by record creators and archivists.[341] Casey has likewise observed that, "all Web sites exist in a physical medium, whether a hard drive, mainframe, CD-ROM, or computer disc. A Web site, after all, is a set of computer files."[342] Casey draws upon the work of Charles B. Lowry, who described the Internet as "a large distributed computing system with a decentralized administration."[343] Websites are made up of files stored on servers, which are computers that contain both server software and the website's component files, including HTML documents, CSS stylesheets, JavaScript files, images, and any

[340] William J. Mitchell, *Placing Words: Symbols, Space, and the City*. Cambridge: MIT Press (2005): 3.
[341] Anne Kenney and Nancy McGovern, "Preservation Risk Management for Web Resources," *Information Management Journal* 36, no. 5 (2002): 52-61.
[342] Casey, "The Cyberarchive," 307.
[343] Charles B. Lowry, "Putting the Pieces Together—Essential Technologies for the Virtual Library," *Journal of Academic Librarianship* 21 (July 1995): 297–300.

other code or digital objects used in the production of the website.[344] The server also includes software that facilitates remote user access to hosted component files; for example, a Hypertext Transfer Protocol (HTTP) server.[345] Users are able to access the website through web browsers on their own home computers by navigating to a URL which requests site files from the web server that hosts them, delivering them to the user's screen. Given this infrastructure, one may question what, precisely, constitutes the archival record. In the previous chapter, personal digital records were stored on floppy disks, hard drives, and complete computers. In some instances, they were accompanied by the software required to read them. To work with the personal records discussed throughout this chapter, archivists have had to reevaluate what it means to collect a record when it is stored on remote, commercially-controlled servers.

When collecting locally-stored personal digital records, as in the previous chapter, archivists relied upon strategies that preserved their material integrity – their fixity – in order to ensure the transfer of authentic, reliable digital information. While the web-based personal records examined in this chapter do exist as data on a fixed medium, in possession of their own fixity, the accessibility of their interface-level data on the open web has allowed for the establishment of archival records that are more transformative than fixed. Web archiving strategies, and specifically web crawling technologies, have been employed in the majority of cases studied in this chapter, resulting in archival records that have been derived from their original, live, web-based versions. Richard Cox has suggested that when collecting personal web-based records that have been created and stored on remote servers, "fixity might not be the goal anymore."[346] Instead, he has proposed,

[344] Mozilla Developer Network, "What is a Web Server?" https://developer.mozilla.org/en-US/docs/Learn/Common_questions/What_is_a_web_server.
[345] Mozilla Developer Network, "What is a Web Server?"
[346] Cox, *Personal Archives and the New Archival Calling*, 251.

"a substantial portion of future work might be the effort to identify personal and family archives on the Web and to try to follow them as they are moved and removed."[347] Abby Smith has summarized the effects of this this shift effectively, writing that:

> "as intellectual content migrates from print, film, and tape to electronic formats, it moves from a world characterized by the fixity and relative permanence of the medium into one in which the stability of the text is easily compromised, the permanence of the intellectual content hard to ensure, and the means of accessing information controlled by the user, not the creator."[348]

Indeed, web archiving strategies, and particularly web crawling, seem to have been widely adopted, at least partially, in response to the perceived ephemerality and transience of web-based records. Web crawling offers archivists a proactive way to intervene at various stages in the life of a website, capturing records before they can be deleted, misplaced, or forgotten, rather than waiting for donors to deposit their websites in the archives.

Only one of the collections studied in this chapter has acquired web-based personal records without turning to web crawling tools. That is the Katie Lee Collection, held by Northern Arizona University's Cline Library.[349] Lee's collection most closely resembles those discussed in the previous chapter in that it is a hybrid collection, containing both digital and non-digital records. Within the collection, a dedicated section, Record Group 6: Websites, is home to the collection's web-based materials, including Lee's personal website, www.katydoodit.com.[350] The website continues to exist as a live WordPress website, hosted through the service A2 Hosting, based in

[347] Cox, *Personal Archives and the New Archival Calling*, 251.
[348] Abby Smith, "Introduction," *Enduring Paradigm, New Opportunities: The Value of the Archival Perspective in the Digital Environment,* 2000. https://www.clir.org/pubs/reports/pub89/.
[349] Katie Lee Collection, 1719-2019, Northern Arizona University, http://www.azarchivesonline.org/xtf/view?docId=ead/nau/lee_katie.xml&doc.view=print;chunk.id=0.
[350] Within archival arrangement and description, a "record group" is a term often seen in organizational collections. The term refers to "a hierarchical division that is sometimes equivalent to provenance, representing all the records of an agency and its subordinate divisions." https://www2.archivists.org/glossary/terms/r/record-group.

Michigan.[351] Hosting services are retained by NAU, rather than by Lee, but the material conditions of the website itself are largely unchanged. To retain and maintain a live website preserves much of the experience of accessing the original artifact, but comes with its own sustainability risks and requirements. To maintain a live website within an archival context requires that archivists make a series of decisions about how they will – or will not – maintain and update the site over time, while being mindful of the potential for making unintentional revisions or changes to the original object. Lee's site exists today in what can be described as "ongoing maintenance," in the meaning of this term that has been articulated in the Socio-Technical Sustainability Roadmap. This refers to a period of time after the active creation and development of the website has ceased, in which "regular, non-transformative activities to sustain the project are undertaken. These activities may include software updates, maintaining hardware and operating systems, updating incorrect or outdated content, or other such behavior."[352] Routine acts of maintenance, such as WordPress software updates or the renewal of hosting services, are used in order to keep Lee's personal website intact.

In each of the other four collections studied in this chapter, archived versions of web-based records have been collected through the use of web crawling. Web crawling employs an automated agent, commonly referred to as a crawler, robot, or spider, which navigates websites and collects data about their contents. The crawler is configured to capture and store web-based records.[353] Web crawlers visit a particular website, or the "seed" of the collection and follow links from the

[351] A Hosting Checker search for www.katydoodit.com indicates that NAU hosts the site through A2 Hosting, Inc., based in Michigan. https://hostingchecker.com/#katydoodit.com.
[352] "Module A2: How Long Do You Want Your Project to Last?" The Socio-Technical Sustainability Roadmap, https://sites.haa.pitt.edu/sustainabilityroadmap/a2-longevity/.
[353] Maria Praetzellis, "Glossary of Archive-It and Web Archiving Terms," Archive-It, https://support.archive-it.org/hc/en-us/articles/208111686-Glossary-of-Archive-It-and-Web-Archiving-Terms.

page, copying content from each of the pages it visits.[354] Web crawling allows for the collection of large quantities of publicly-accessible web-based records without requiring access to back-end databases or site files. The versions that reside in archives are not fixed representations of complete, live websites, but are instead representations of each site's surface-level interface, which are often referred to as "snapshots" – not a copy of the complete website, but rather a representation of what the website would have looked like to a user at the time the snapshot was taken. In other words, snapshots are dynamic pictures of the website's content and structure. They represent a fixed version of the website as it appeared at the time of capture. Should the live version of the website disappear or change, the archived snapshot can provide continued access to a stable copy of a previous version of the site.

During its time as an active service, GeoCities offered users both web hosting and a suite of editing tools to support website creation without any pre-existing knowledge of HTML or CSS. As a live service, GeoCities stored its data on a "cluster of 65 Sun Microsystems Ultra 2 ad Enterprise 5000 machines, running Sun's Solaris operating system and the public-domain Apache Web software."[355] These components of GeoCities' material infrastructure were located in the northern California town of Santa Clara, while staff managed them remotely from offices in Marina del Rey and Santa Monica. Necessary maintenance that had to be conducted locally in Santa Clara, "such as swapping cards and replacing disks," was outsourced to local contractors.[356] The personal websites created and hosted within the GeoCities platform relied upon this complex infrastructure, which was largely invisible to its users.

[354] "What is Web Archiving? History, Technology, Collections," *National Records of Scotland Open Book blog*, November 24, 2017. https://blog.nrscotland.gov.uk/2017/11/24/what-is-web-archiving-history-technology-collections/.
[355] Janah, "GeoCities Copes with Rapid Growth," *InformationWeek*, December 7, 1998.
[356] Janah, "GeoCities Copes with Rapid Growth."

In April 2009, Yahoo, which had acquired GeoCities ten years earlier, announced that it would cease to support the service, and would be permanently deleting all extant GeoCities websites by October of the same year.[357] Efforts to collect GeoCities websites, so as not to permanently lose access to these records, began in earnest. Among these efforts were those spearheaded by the Internet Archive, which created a GeoCities special collection with its Wayback Machine service, and Archive Team, an affiliation of self-described "rogue archivists." Archive Team's collection is among the most prolific, containing nearly one terabyte worth of public-facing GeoCities websites.[358] In a presentation about the GeoCities collection at the 2011 Personal Digital Archiving conference, Jason Scott of Archive Team reported that the collection, which consisted of snapshots of 900 gigabytes worth of GeoCities websites, was as large a portion of GeoCities' live content as could be collected in the six months between notification and deletion; however, Scott noted, it was unknown how many GeoCities websites had not been captured and were permanently lost.[359] The question of extent is not unique to GeoCities. As has been observed elsewhere, "there has never been any way to enumerate all web pages; so, all attempts to archive the web are to some extent incomplete."[360]

[357] Leena Rao, "Yahoo Quietly Pulls the Plug on GeoCities," *TechCrunch* April 23, 2009, https://techcrunch.com/2009/04/23/yahoo-quietly-pulls-the-plug-on-geocities/.
[358] "About," One Terabyte of Kilobyte Age, https://blog.geocities.institute/about.
[359] Scott, "The Splendiferous Story of Archive Team."
[360] Scott A. Hale, Grant Blake, and Victoria D. Alexander, "Live versus Archive: Comparing a Web Archive to a Population of Web Pages," in *The Web as History* ed. Niels Brügger and Ralph Schroeder (London: UCL Press, 2017), 46.

Figure 6 Hard drive containing a copy of the Archive Team's GeoCities collection.

The 900 gigabytes of GeoCities collected by Archive Team were subsequently compressed, resulting in a 645 gigabyte file that the Team released as a torrent in 2011, which remains available for download on the torrent site The Pirate Bay.[361] Describing the collection, Scott noted that users could download the entirety of the remains of GeoCities and store it "on a hard drive the size of a pack of cards."[362] Indeed, my own copy of the Archive Team's GeoCities collection, downloaded from the Archive Team's torrent, is stored on such a hard drive (Figure 6).

[361] The Pirate Bay originated in Sweden in 2003, and supports peer-to-peer torrent downloads of entertainment media and software, https://thepiratebay.org/index.html. The site has periodically been at the center of censorship and copyright infringement controversies. See "Pirate Bay Hit with Legal Action," *BBC News* January 31, 2008, http://news.bbc.co.uk/2/hi/technology/7219802.stm; Elizabeth Barber, "Pirate Bay Goes Offline After a Raid in Sweden," *Time*, December 10, 2014, https://time.com/3627330/pirate-bay-offline-raid-sweden/; Samuel Gibbs, "Pirate Bay Revived by Rival Piracy Site," *The Guardian*, December 15, 2014, https://www.theguardian.com/technology/2014/dec/15/pirate-bay-revived-by-rival-piracy-site-torrent-isohunt.
[362] Scott, "The Splendiferous Story of Archive Team."

However, as Dragan Espenschied elaborated in the project's GitHub README file, users of the archive will "need a lot of hard drive space to seriously work with GeoCities."[363] Espenschied recommended "setting up an LVM [Logical Volume Manager] spanning two 2TB disks" in order to navigate and analyze the collection more easily.[364] The technical challenges of managing and using this collection present a notable barrier to access. Other entities, unaffiliated with Archive Team, have mounted their copies of the downloaded collection online, attempting to provide additional points of access. The result of Archive Team's method of storage and dissemination is a network of locally-stored GeoCities collections, including my own. In some instances, data from the Archive Team's collection has been used in sites mirroring the GeoCities archive in order to support web-based access to the collection. The Archive Team website directs users to several of these, including GeoCities Archive Project, Reocities, Oocities, Internet Archaeology, and the now-defunct geocities.ws, noting that each of these initiatives had attempted their own capture of live GeoCities sites, and that "all groups appear to have gotten different amounts of the GeoCities collection, and most are now sharing data to track down gaps and share copies."[365] Archive Team also points to the web-based GeoCities collection maintained by the Internet Archive, with whom Archive Team has collaborated in establishing this collection.

The Internet Archive's GeoCities Special Collection takes the form of a series of snapshots of live GeoCities sites, captured by the Wayback Machine and accessible through its primary search interface.[366] The contents of the GeoCities Special Collection are stored and maintained

[363] Dragan Espenschied, "README," GeoCities, https://github.com/despens/Geocities/blob/master/README.
[364] Espenschied, "README."
[365] "The GeoCities Project and Friends," Archive Team, https://www.archiveteam.org/index.php?title=GeoCities#The_GeoCities_Project_and_Friends.
[366] GeoCities Special Collection 2009, Internet Archive, https://archive.org/web/geocities.php.

within the Internet Archive's overarching data storage infrastructure. The WARC files generated through web crawling are located within the Wayback Machine's overarching infrastructure.[367]

These snapshots portray the public-facing interface layer of websites as they existed at the moment of capture. They are transformative, saving a dynamic live website as a stable WARC format, flattening the website's component files into an archival package. The Internet Archive relies upon a network of crawlers, primarily Heritrix and Alexa Internet, to crawl and capture public-facing websites into files that adhere to the ISO 28500:2009 Web ARChive file format standard.

While web crawlers are designed to capture the surface, or interface, layer of a website, they are unable to capture certain website elements. Archived snapshots of live websites often feature broken links, indicating points where crawling was unable to capture content. Many web crawlers are specifically unable to capture dynamic content, including "[websites] generated by a user search, JavaScript, drop-down menus, and streamed media such as embedded YouTube videos."[368] Searching through the Internet Archive's GeoCities collection, one encounters many examples of websites with broken links where media files were once embedded. The broken image links on the personal page "Scarlet Enterprises Home Page," for example, once incorporated GIF files, hosted elsewhere, within the page (Figure 7).[369]

[367] Mat Kelly, Michele C. Weigle, "WARCreate: Create Wayback-Consumable WARC Files from Any Webpage," *JCDL '12: Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries* (June 2012): 437. For WARC specifications, see "WARC, Web ARChive file format," https://www.loc.gov/preservation/digital/formats/fdd/fdd000236.shtml.

[368] "What is Web Archiving? History, Technology, Collections," *National Records of Scotland Open Book blog*, November 24, 2017.

[369] The Graphics Interchange Format (GIF) was a still image file format frequently used in GeoCities websites. See http://www.loc.gov/preservation/digital/formats/fdd/fdd000133.shtml for format specifications. GIFs were such a hallmark of GeoCities that the Internet Archive developed another special project, GifCities, to provide ongoing access to the animated GIFs that so many people associate with that service. See https://gifcities.org/ for the searchable GIF database.

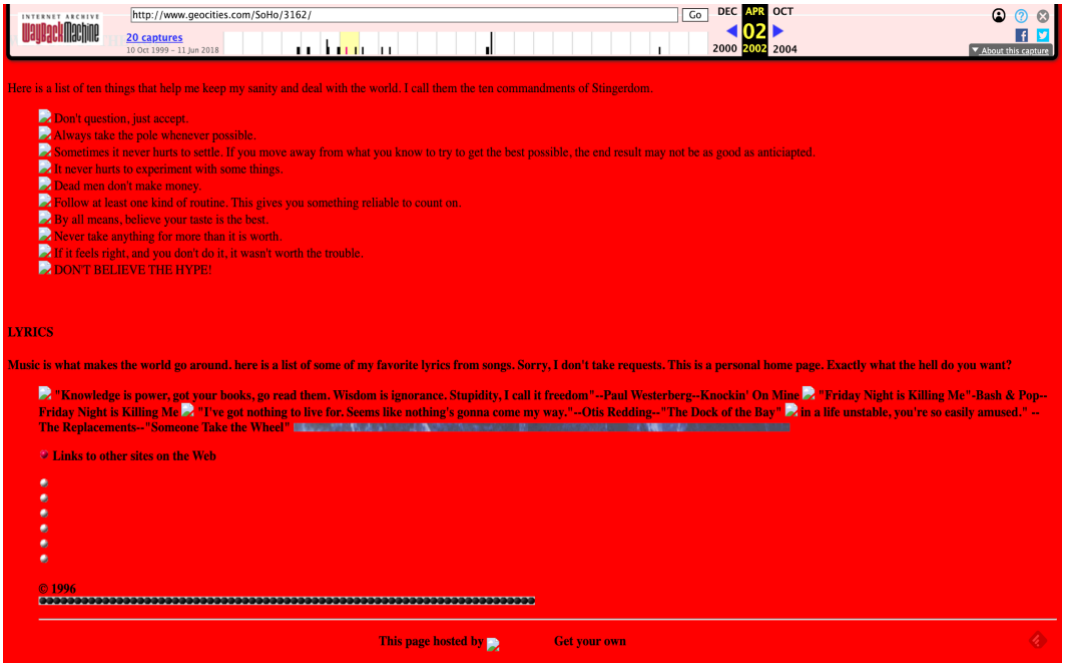

Figure 7 Archived version of Scarlet Enterprises Home Page.

The born-digital records in the Mormon Missionary Collection held by Brigham Young University also rely upon the Internet Archive's web crawling and data storage infrastructure. BYU uses the Internet Archive's Archive-It tool to capture and preserve born-digital, web-based records for this collection, which consists primarily of the blogs maintained by missionaries to document their experiences of service.[370] The majority of these digital records were created with free software from Blogger and WordPress, and have been captured and made accessible to the public with Archive-It. These archived versions of the live blogs are stored as WARC files, hosted by the Internet Archive and stored in their data centers.[371] In studying these collections, it has become clear that the Internet Archive plays a central role, not only through its own collections,

370 Mormon Missionary Collections, https://lib.byu.edu/collections/mormon-missionary-collections/.
371 "Learn More," Archive-It, https://archive-it.org/blog/learn-more/.

but through the provision of web crawling tools and services, including Archive-It, and data storage services for cultural heritage institutions.

The Internet Archive maintains data centers in three locations in California: Redwood City, San Francisco. Data are stored with the Petabox storage system, a custom solution designed to house large quantities of data at low cost and low energy usage.[372] In 2010, each Petabox contained "240 2-terabyte disks in 4U high rack mounts," with 8 Petabox units totaling approximately 4 million gigabytes of data.[373] By 2014, the Internet Archive's storage consisted of "4 data centers, 550 notes, 20,000 spinning disks," with the Wayback Machine's data totaling 9.6 PetaBytes.[374] The totality of the Archive's data storage has been mirrored in both Egypt and Amsterdam. In 2011, Brewster Kahle described the geographic dispersion of storage locations, stating that "our earthquake zone is backed up in the turbulent Mideast and a flood zone. I won't sleep well until there are five or six backup sites."[375] Kahle's comment is a stark reminder that web-based records are not only material, but are closely connected to and affected by the geographic, social, and political environments in which they are situated.

Kenney and McGovern have further considered data preservation from this perspective, advocating for a risk management approach to sustaining web-based digital resources. Their research suggests that the greatest threats to these records are not "how well the site is maintained or even how often it is backed up but whether the backup tapes are stored in the same room as the server - increasing the chance that a single catastrophic event could destroy them both."[376]

[372] "Petabox," Internet Archive," https://archive.org/web/petabox.php.
[373] Jeff Kaplan, "The Fourth Generation Petabox," *Internet Archive Blogs*, July 27, 2010, https://blog.archive.org/2010/07/27/the-fourth-generation-petabox/.
[374] "Petabox," Internet Archive," https://archive.org/web/petabox.php.
[375] Brewster Kahle, "Universal Access to All Knowledge," *The Long Now Foundation*, November 30, 2011, http://longnow.org/seminars/02011/nov/30/universal-access-all-knowledge/.
[376] Kenney and McGovern, "Risk Management for Web Resources."

Concerns about the individual control that is relinquished in remote data storage infrastructures underscore the ways in which web-based records read as both immaterial (accessible "virtually" anywhere) and material (stored on physical, geographically situated media). While the bit-level concept of fixity is not, as Cox and Smith have suggested, as central a concern in web archiving as it has been in the preservation of other born-digital records, there is a larger-scale concern about the lack of consistent, or fixed, access to these materials.

The infrastructure of the World Wide Web has a clear impact on archival practice. Websites are stored remotely, within material and geographically situated data centers beyond the control of the individuals who create them – and, often, the archivists who collect them. Concerns about the loss of digital cultural heritage as a result of the short lifespans of many websites have led to more contemporaneous collecting practices, including the use of web crawling tools to capture of websites in real time, and the preservation of those snapshots. Web crawling is transformative, producing an archival record that takes a different format than the original it represents. The following section explores how these methods are used to establish control over web-based personal records.

5.3 Custody and Control

Concerns about the potential loss of web-based records, whether as a result of deletion, link rot, or technological failure, have been a call to action for archivists. As several of the collections explored throughout this chapter demonstrate, however, the pressure to quickly collect personal records from the open web before they are lost or destroyed raises a number of complicated issues. In particular, it is unclear who owns and controls the records created and stored

within services like GeoCities, Blogger, and WordPress.[377] Without a clear understanding of ownership, it has proven challenging for archivists to establish consistent collecting policies, particularly with regard to acquiring custody of personal records. Both legal custody and physical custody are complicated by the infrastructure of the World Wide Web. The collections explored throughout this chapter reveal a range of questions about how web-based personal records are controlled, and by whom, and how these conditions impact archival accessioning practices.

The question of who owns web-based personal records can be difficult to answer, for both archivists and individual record creators themselves. As Catherine C. Marshall, Frank McCown, and Michael L. Nelson have observed in their research on personal archiving strategies for web-based materials, many individuals have – or believe that they have – less control over their web-based records, such as websites or blogs, than they do over their locally-stored digital records.[378] This uncertainty stems in part from the fact that these records are often stored remotely, on servers controlled by commercial service providers or other external institutions.[379] However, the question of ownership is often more complicated than it appears, with both individual record creators and service providers exerting varying degrees of control over records. When individuals store their records with service providers, the records are not only remotely-stored, they are subject to the terms and conditions of the service provider. The open question of the legal ownership of web-based personal records results in uncertainty as to who may decide to collect them, and through

[377] This chapter considers websites that have been created with the open source website publishing software WordPress (www.wordpress.org), which is "owned by no one individual company" and can be hosted anywhere, as well as those created with the hosted blogging service WordPress (www.wordpress.com) which is operated by the company Automattic. "WordPress vs WordPress.com," WordPress Support, https://wordpress.org/support/article/wordpress-vs-wordpress-com/. Katie Lee's personal website was created with the former, and a number of the blogs in the Mormon Missionaries Collection were created with the latter.
[378] Catherine C. Marshall, Frank McCown, and Michael L. Nelson, "Evaluating Personal Archiving Strategies for Internet-Based Information," Proceedings of Archiving 2007 Arlington, VA, (May 21-24, 2007): 151.
[379] Marshall, McCown, and Nelson, "Evaluating Personal Archiving Strategies for Internet-Based Information," 151.

what means. Kenney and McGovern have explored these issues, recommending a three-pronged approach: "collaborating with publishers to preserve licensed content, developing policies and guidelines for creating and maintaining Web sites, and assuming archival custody for Web resources of interest."[380] The collections explored within this chapter reveal a range of approaches to acquiring custody, both legally and physically, of web-based personal records, working with a variety of the stakeholders identified by Kenney and McGovern and, in some instances, without the input of any of them.

Table 6 Methods of acquiring web-based personal digital records

Collection	Collecting Institution	Acquisition Method	Acquisition Source
GeoCities Collection	Archive Team	Web Crawling	World Wide Web
GeoCities Special Collection	Internet Archive	Web Crawling	World Wide Web
Katie Lee Collection	Northern Arizona University	Transfer of Hosting Service	Record Creator (Katie Lee)
Mormon Missionaries Collection	Brigham Young University	Web Crawling	World Wide Web
Zine Web Archive	Library of Congress	Web Crawling	World Wide Web

The infrastructure of the open web has allowed archivists to take personal records into their custody in new ways. Specifically, decentralized, public access to records on the open web has supported departures from traditional means of acquiring personal records. As Table 3 shows, only one of the collections discussed in this chapter, the Katie Lee Collection at Northern Arizona

[380] Kenney and McGovern, "Preservation Risk Management for Web Resources."

University, was acquired through a formal transfer of custody, and specifically, a transfer of hosting services, for the complete, live website.[381]

Figure 8 Screenshot of Katie Lee's personal website.

In this respect, the Katie Lee Collection stands as an example taking custody of records that most closely resembles the practices described in Chapter Four, in which records are transferred directly from creator to archives. When one navigates to Katie Lee's personal website at www.katydoodit.com, they will encounter a statement on the homepage that informs them of Lee's death on November 1, 2017 (see Figure 8). "Prior to her passing," the statement reads, Lee "made arrangements with Northern Arizona University's Cline Library to care for and provide

[381] "Record Group 6," Katie Lee Collection, http://www.azarchivesonline.org/xtf/view?docId=ead/nau/lee_katie.xml&doc.view=print;chunk.id=0.

access to her archival legacy."[382] The Katie Lee Collection finding aid confirms the presence of this website among Lee's other archival materials.

The bulk of Lee's materials, including digital and non-digital records, were received by NAU in 2008, with subsequent deposits made in 2011, 2014, 2015, and 2017.[383] It was in the final 2017 accession that Lee's personal website came into archival custody.[384] Of the five collections examined in this chapter, Lee's is the only one that has been acquired by an archive through a direct transfer of legal custody from the creator. In this respect, it most closely resembles the collections discussed in the previous chapter. Though the conditions of remote storage impact the transfer of physical custody, the transfer of legal custody and control from creator to archives aligns with traditional archival practices.[385] By assuming responsibility for the hosting services of the website, NAU took control of the site and its contents, including both public-facing front-end and administrative back-end files.

The other four collections studied in this chapter were acquired directly from the open web, with the use of web crawling tools. Web crawling enables the collection of records (or more accurately, as discussed in the preceding section on Materiality, of transformed, stable versions of those records) directly from the open web, without the legal transfer of custody from record creator to archival repository.[386] In this respect, web crawling as a means of acquisition can be understood as a dramatic shift in practice. From a technological standpoint, the consent of creators is not required for acquisition; the imperative to involve creators must instead be shaped by the policies

[382] Home, Katie Lee, https://www.katydoodit.com/. Accessed May 13, 2020.
[383] "Administrative Information," Katie Lee Collection.
[384] Katie Lee Collection, Record Group 6.
[385] As verified by a Hosting Checker search for www.katydoodit.com, NAU does not host the site locally, but instead uses A2 Hosting, Inc., based in Michigan. https://hostingchecker.com/#katydoodit.com. NAU is responsible for maintaining the web hosting services and administering the website.
[386] Niu, "An Overview of Web Archiving."

and ethical frameworks in place within each collecting institution. The Mormon Missionaries Collection, Zine Web Archive, and each GeoCities collection take different approaches to acquiring custody of the crawled versions of web-based personal records.

The Library of Congress's Zine Web Archive was created to supplement their collection of physical zines, and is situated within the Library's general Web Archiving program.[387] The websites included in the Zine Web Archive have been identified for collection by the Library's Recommending Officers and Subject Specialists, but the collection website notes that "this collection is primarily curated by the Collection Specialist for Women's, Gender, and LGBTQ+ Studies," as these are areas of particular interest according to the collection's scope and content description.[388] Websites are captured with web crawlers, primarily Heritrix, on a recurring basis, with the majority of the sites being captured monthly or yearly, and with a smaller selection of sites "targeted for capture quarterly, twice-yearly, or once."[389] The Library's Web Archiving program makes an effort to include website creators in the collecting process, explaining that they send an email to notify site owners that they would like to include the site in their collections.[390] Within that email, site owners can access a permissions tool that grants or denies permission for the Library to capture their website; these responses are then recorded in a database and used to identify and track sites to be crawled.[391] Abbie Grotke, the Library's Web Archiving Team Lead, has noted that taking a permission-based approach to web archiving poses challenges that have the

[387] The primary Library of Congress Web Archiving site can be accessed at https://www.loc.gov/programs/web-archiving/about-this-program/. A guide to the Zine Collections at the Library of Congress is located at https://guides.loc.gov/zines/collections.
[388] "About this Collection," Library of Congress Zine Web Archive, https://www.loc.gov/collections/zine-web-archive/about-this-collection/.
[389] "About this Collection," Library of Congress Zine Web Archive.
[390] "For Site Owners," Library of Congress Web Archiving, https://www.loc.gov/programs/web-archiving/for-site-owners/.
[391] "Frequently Asked Questions For Site Owners," Library of Congress Web Archiving, https://www.loc.gov/programs/web-archiving/for-site-owners/frequently-asked-questions/.

potential to shape the collections in meaningful ways. A Legal Issues Roundtable held during the 2012 International Internet Preservation Consortium generated a list of common challenges to permission-based web archiving, including the "lack of response from site owners. Members seeking permission reported a 30-50% response rate; it's not that websites are denying permission. They just aren't responding to our attempts to contact them."[392] Members noted that as a result, collections were often "patchy" or "unbalanced."[393] Further, the "tremendous effort required to contact site owners and notify or obtain permission can sometimes overwhelm staff resources."[394] Other collecting institutions have opted for strategies that are not permission-based; instead, they may collect widely, often without notifying individual record creators, occasionally providing mechanisms through which creators may request to have their records removed from the archive.

Like the Zine Web Archive, the web archiving portion of the Mormon Missionary Collection is an extension of a non-digital collection.[395] BYU uses Archive-It to crawl blogs that were created and maintained by missionaries during their periods of service; the collection relies upon a multi-pronged approach to acquisition. In addition to crawling blogs that employees of BYU's library and archives have identified for acquisition, users of the collection are encouraged users to submit links for missionary blogs they would like to see included in the web archive; users do not have to have created the blogs in order to suggest them for inclusion. Indeed, individual

[392] Abbie Grotke, "Legal Issues in Web Archiving," *The Signal*, May 30, 2012, https://blogs.loc.gov/thesignal/2012/05/legal-issues-in-web-archiving/.
[393] Grotke, "Legal Issues in Web Archiving."
[394] Grotke, "Legal Issues in Web Archiving."
[395] See the Mormon Missionary Diaries, https://lib.byu.edu/collections/mormon-missionary-diaries/. A portion of this collection has been digitized and made available online. Though personal records that have been created in analog format and subsequently digitized are beyond the scope of this dissertation, it should be noted that they are subject to many of the same ethical issues present in born-digital personal archives. Tara Robertson's keynote address at the 2016 Library and Information Technology Association Forum, "Not All Information Wants to Be Free," provides a useful introduction to the issues that arise specifically in digitized collections, https://tararobertson.ca/2016/lita-keynote/.

record creators do not need to be involved in, or even notified about, the inclusion of their blogs and websites in the web archive. Instead, BYU offers an opt-out mechanism for creators who may discover that their blogs have been included in the collection and wish to have them removed. Their website states that "if you are an owner of content that has been harvested by the Web Archive and wish your material not be included in the Web Archive please contact the Copyright Licensing Office.[396] The terms of such a removal request will be discussed in further detail in this chapter's Privacy and Publicity section.

While the ability to crawl selected public-facing websites for archival purposes without the formal permission of record creators has led to a range of collecting practices that differ from those used to acquire locally-stored personal records, the efforts to preserve websites created in the GeoCities platform constitute what is perhaps the most dramatic departure from traditional approaches to acquiring personal records of all of the collections explored within this chapter. Each of the GeoCities collections serves as an example of collecting personal records at the level of the service provider, rather than acquiring a collection on the basis of its authorship, as in the Katie Lee Collection, or even based on its specific subject matter, as in the Mormon Missionary Collection or Zine Web Archive.

The decision to collect all extant GeoCities websites at scale is often attributed to the sociotechnical context of their creation and subsequent destruction, rather than because of the specific individuals who created them. The collections are often described as documentation or evidence of early web history. The explanation given on the Archive Team website provides an example of a typical justification given for collecting the sites *en masse*:

[396] Brigham Young University Archive-It Home, https://archive-it.org/home/byu.

> "While the natural urge by some would be to let GeoCities sink into obscurity and death, leaving nothing in its wake but bad memories and shudders of recognition at endless 'under construction' GIFs, the fact remains that GeoCities was for millions of people the first experience dealing with the low-cost, full-color, world-accessible website and all the possibilities it contained. To not at least have the option of browsing these old sites would be a loss of the very history of the web from the side of the people who came to know it, not the designers who descended upon it."[397]

Under the mission statement of preserving early web history, the efforts to capture GeoCities websites before they were deleted by Yahoo are typically positioned as attempts to rescue digital records before they are lost permanently.[398] This framing simultaneously obfuscates the roles and intentions of individual website creators and justifies the rapid collection of all websites from a platform without creator involvement. In the narrative surrounding the capture of GeoCities, the key players are the service provider preparing to destroy the records and the archivists and volunteers who rescued them from destruction.

Like BYU, the Internet Archive involves individual website creators in the long-term control of the collection by providing a system through which users may request that their websites be removed from the Wayback Machine; users can email the Internet Archive about the website in question in order to begin a review process.[399] Archive Team's GeoCities collection, however, is not so easily altered. Because Archive Team has made the collection accessible by way of a torrent download of the complete collection, it is not possible to control any versions of the collection that have been downloaded. The result of this mode of dissemination is a system of

397 "GeoCities - Glorious History," Archive Team, https://www.archiveteam.org/index.php?title=GeoCities.
398 Both Archive Team and the Internet Archive frame their collections in this way. Press coverage likewise reinforced the notion of the collections as a data rescue. See Olia Lialina, "GeoCities' Afterlife and Web History," One Terabyte of Kilobyte Age, July 30, 2017, https://blog.geocities.institute/archives/6418; Scott Gilbertson, "GeoCities Lives on as a Massive Torrent Download," *Wired*, November 1, 2010, https://www.wired.com/2010/11/geocities-lives-on-as-massive-torrent-download/; Rieke Jordan, "In the Ruins of GeoCities," *Response Journal* 1, no. 1 (2016), https://responsejournal.net/issue/2016-11/article/ruins-geocities.
399 "Using the Wayback Machine," Internet Archive, https://help.archive.org/hc/en-us/articles/360004651732-Using-The-Wayback-Machine.

geographically dispersed copies of the complete collection, individually controlled by those who have downloaded them. Therefore, individuals who may wish to have their records excluded from the archive have no recourse.

Each of the approaches to taking personal, web-based records – or the snapshots of them captured with web crawling – described in the preceding pages has certain privacy implications. Without the involvement of record creators or the establishment of a formal agreement about the terms under which these materials will be made accessible, treatments of privacy are left almost entirely to the discretion of the archivists who control these materials. The following section explores these issues in greater detail.

5.4 Privacy and Publicity

Archivists who collect personal records have long endeavored to balance the need to protect the privacy of record creators and subjects with the duty of providing access to the users of archival collections. The concepts of privacy and publicity, as they have traditionally been understood in personal archives, have been further complicated by the infrastructure of the World Wide Web. As both Richard Cox and James Rule have argued, emerging web-based information technologies blur the line between public and private in a multitude of ways, many of which are relevant to the work of archivists.[400] Sara Hodson has likewise argued that, "the archival privacy issues continue in the online age and, in fact, they can be more challenging than ever, due to… the ubiquity of the Internet, and the ease with which data can be collected, posted, altered, and widely

[400] Richard J. Cox, *Personal Archives and a New Archival Calling,* 110; James B. Rule, *Privacy in Peril* (New York: Oxford University Press, 2007): 113.

disseminated."[401] Each of the collections discussed in this chapter contains personal records that were originally publicly accessible on the open web. The fact that these websites and blogs were publicly accessible on the open web has had an impact on approaches that archivists have taken to both collecting those records and the ways that they subsequently make them accessible.

As discussed in the previous section on Custody and Control, web crawling strategies allow archivists to collect personal records from the open web, without requiring the involvement of individual record creators. Privacy is particularly complicated in this particular context because, as many have argued, websites are, most often, already openly available to the public, making it more difficult to determine what information might be considered private, and in need of protection.[402] The collections examined in this chapter reveal a range of approaches that have been taken to involving record creators not only in the acquisition process, but in making decisions about how the archived versions of these records will be made accessible to users. Though all of these records were once found online, their archived versions may be made accessible in a variety of ways. Some collections, such as the Katie Lee web-based materials, the Mormon Missionary Collection, and the Internet Archive's GeoCities Special Collection have remained online. Katie Lee's personal website is unique among these in that it has remained intact as a live website, hosted by NAU as the result of a formal agreement with Lee and her estate.[403] Others, such as the Zine Web Archive, have employed a hybrid approach, in which record creators may decide whether or not the archived versions of their websites will be made accessible online. The Archive Team's GeoCities collection is unique among these collections in that has not made its contents browsable

[401] Sara Hodson, "Archives on the Web: Unlocking Collections While Safeguarding Privacy," *First Monday* 11, no. 8 (2006), https://firstmonday.org/ojs/index.php/fm/article/view/1389.
[402] Hou, "Concept, Process, and Principle: Research on the Privacy Protection in Web Archiving," 110.
[403] "Record Group 6," Katie Lee Collection, Northern Arizona University.

or searchable online, but instead has made them accessible through a torrent download of the entire collection. This section will describe the privacy implications of these modes of access.

Most of these collections can be described as having taken an "opt-in" or an "opt-out" approach to privacy. Lauren Kaufman has described these as two privacy "regimes" that are commonly used to determine a user's "consent status in relation to a specific data collector."[404] In a web archiving context, an opt-in regime requires that record creators explicitly consent to including their records in an archival collection; an opt-out regime collects records without prior consent from record creators, and instead requires them to make a request to the collecting institution if they wish to be excluded from the archives.

The Library of Congress' Zine Web Archive employs an opt-in approach to collection development. The Library's web archiving program has a documented policy of asking record creators for permission to make archived websites accessible online prior to collecting them.[405] Additionally, at the creator's request, access to archived versions of websites in the collection can be limited to local computers within the Library's reading rooms.[406] As reported in the previous section, the Library contacts website owners via email, an approach that they have noted typically garners a response rate of under 50%. Though this approach may result in fewer acquisitions, it involves creators in the process and helps to minimize potential violations of privacy.

The Internet Archive's GeoCities Special Collection and BYU's Mormon Missionary Collection both employ an opt-out approach in their web archiving efforts. While BYU has provided a mechanism through which users may submit blogs for inclusion in the collection, it

[404] Lauren Kaufman, "To Opt-In or Opt-Out?" *Popular Privacy*, March 6, 2020, https://medium.com/popular-privacy/to-opt-in-or-opt-out-5f14a10bae24.
[405] "For Site Owners," Library of Congress Web Archiving, https://www.loc.gov/programs/web-archiving/for-site-owners/.
[406] "For Site Owners," Library of Congress Web Archiving.

does not require that those users by the creators of the blogs in question. BYU's primary Archive-It page notes that the materials included in these collections may be protected by copyright, and users who wish to use any collection materials for purposes other than personal research must obtain permission from the copyright holders.[407] If blog creators should wish to remove their materials from the collection, they must contact the university's Copyright Licensing Office in order to initiate a removal process. [408] The GeoCities Special Collection likewise collects sites without first seeking permission from creators, and provides an email address that website creators may write to in order to request the removal of a website.[409] The opt-out approach garners a more comprehensive collection, but it denies agency to record creators. Further, without a notification system, users must first learn that their records are included in the collection in order to request that it be removed.

Archive Team's GeoCities collection was also established without the involvement or consent of individual record creators; however, it provides no "opt-out" mechanism. is unique in that its contents cannot be browsed online. Instead, users must download the complete collection torrent in order to access its contents. This presents some barriers to access for those who do not have the storage space or technological confidence to download and navigate the complete collection. It also prohibits website creators from requesting that their websites be removed from the collection: even if Archive team removed the website in question and re-released the torrent without it, there remain an unknown number of distributed copies of the previous version of the collection, stored and individually controlled by those who downloaded it before its re-release.

[407] Brigham Young University Archive-It Collections, https://archive-it.org/home/byu.
[408] Brigham Young University Archive-It Collections.
[409] "Using the Wayback Machine," Internet Archive, https://help.archive.org/hc/en-us/articles/360004651732-Using-The-Wayback-Machine.

These collections demonstrate a range of potential approaches to considering issues of privacy in web-based personal archives. In the cases where archivists have not worked directly with record creators – in other words, in the collections that use an opt-out approach – contextual integrity becomes a particularly useful framework for helping archivists to better understand the impact of their collections. Does collecting and preserving a website or blog that is already accessible on the World Wide Web constitute a violation of privacy? Nissenbaum argues that:

> "it is crucial to know the context - who is gathering the information, who is analyzing it, who is disseminating it and to whom, the nature of the information, the relationships among the various parties, and even larger institutional and social circumstances,"

in order to establish how privacy might be understood.[410] In order to determine whether an action, such as collecting records with the aid of a web crawler, constitutes a violation of privacy is dependent upon a number of variables, including but not limited to "the nature of the information in relation to that context, the roles of agents receiving information, their relationships to information subjects, on what terms the information is shared by the subject, and the terms of further dissemination."[411] These questions are well-suited to archival work, which is fundamentally concerned with context.

Framing privacy as contextual integrity brings more nuance to the discussion around collecting web-based records, as it moves beyond a conception of a personal website as "already public," and instead prompts us to ask what is changed when the website enters into archival custody, is situated in a new space (both physically and intellectually, and is made accessible to new audiences. Importantly, it also takes into account the terms on which the creator or subject

[410] Nissenbaum, "Privacy as Contextual Integrity," 137.
[411] Nissenbaum, "Privacy as Contextual Integrity," 137-138.

has shared the records. These terms are more evident when the creator is also the donor; they become less clear when personal records are collected from the open web. They also require archivists to acknowledge, and perhaps even center themselves, within the complex system of decisions and actions involved in collecting, preserving, and providing access to records. Though Jenkinson and his followers opposed the notion of archivists as active collectors, exerting influence on the documents of history, the collections discussed throughout this chapter demonstrate the active role that archivists have taken in shaping web-based personal archives.[412]

5.5 Conclusion

In 2008, Richard Cox wrote that "it is difficult, at this point, to know just what the impact of the Internet has been on the sense of the personal archive."[413] That the World Wide Web, and specifically services like GeoCities, Blogger, and WordPress, that have allowed individuals to easily create and share records online, has been celebrated by archivists as a valuable contribution to cultural heritage. By collecting these records, archivists have argued that they are able to preserve the records not only of those in power, but of diverse range of experiences and expressions.

Studying collections that contain records from the early years of the web, it is possible to see some aspects of that impact come further into focus. The collections examined within this chapter have revealed significant shifts from the practices applied to personal records that have

412 As discussed in Chapter Three (see page 45), Jenkinson took a firm stance against active collecting in archives, arguing that archives should receive records from the organization within which they were situated. This was a more passive, and less interpretive, approach.

413 Cox, *Personal Archives and a New Archival Calling*, 177.

been acquired on physical media. In particular, the public accessibility of personal websites on the open web has supported new modes of acquisition that raise both practical and ethical questions for archivists.

In conducting this research, it has become evident that, in myriad ways, the Internet Archive is a key player in web archiving, both within and beyond its own institutional bounds. Archivists and information professionals engaged in web archiving, as well as users of web archives, can now recognize not only the impact of the internet on archival practice, but the impact of the Internet Archive, as a service provider that has risen to the level of infrastructure. The WARC file format has become the de facto standard for preserving crawled websites, and the Heritrix web crawler, Wayback Machine, and Archive-It popular tools for capturing and providing access to those preserved files. Both the tools and the comprehensive collecting mission have been impactful beyond the Internet Archive's own collections.

Specifically, the development of web archiving tools, and specifically web crawling, have had a profound impact on archival practice. Of the five collections examined in this chapter, only one – the Katie Lee collection – did not rely web archiving strategies, but instead contained live websites. Web crawling as a web archiving strategy is transformative, producing a record that differs from its live counterpart in both its format and in the information it conveys. Web archiving saves websites as "snapshots," transforming a live, changeable object into one that is stable and fixed – as the name suggests, a "snapshot" of the record as it existed at the time of capture.[414]

Additionally, with the aid of web archiving tools and services, it has become possible for archivists to collect and preserve personal records without the knowledge or consent of their creators. This capability has further complicated the longstanding tension between protecting the

[414] Niu, "An Overview of Web Archiving."

privacy of record creators and subjects, and serving the research needs of researchers.[415] Wielding these tools, archivists have had to determine whether they will collect to the extent that technology permits, or apply more rigorous collecting policies. This chapter revealed that collections may exist within "opt-in" or "opt-out" regimes. The Library of Congress employs an opt-in approach, requesting permission from website owners before archiving their websites with Heritrix crawlers; further, the Library asks website owners to determine whether access to the archived versions of their websites will be made available online, or limited to the Library's on-site reading rooms. The Mormon Missionary Collection at BYU, and both GeoCities collections have employed opt-out approaches, collecting broadly and providing mechanisms through which website owners may request that the archived versions of their materials be removed from the collections. Opt-out approaches may be desirable because they allow for more responsive, contemporaneous collecting practices, which advocates like Kahle have argued are necessary if we are to save web-based records before they are lost.[416] Indeed, as Grotke noted, the more conservative opt-in approach employed by the Library of Congress and many other members of the IIPC result in a response rate of between 30% and 50%.[417] The collections explored throughout this chapter illustrate the persistent disagreement in the field about whether to privilege the record creators or the potential users of archives when developing collecting policies for web archives. Further, they suggest a need for guidance that is driven by theory and policy, rather than by technology.

The impulse to collect broadly and quickly has been driven, in part, by pervasive concerns about the instability of records on the web. Personal web-based records are often stored remotely, stored on devices that are geographically situated far from individual record creators – and

[415] Hodson, "Archives on the Web."
[416] Kahle, "Preserving the Internet."
[417] Grotke, "Legal Issues in Web Archiving."

archivists. These conditions cede control to commercial service providers. Though concerns about technological obsolescence, as evidenced in the previous chapter, still persist, this research has suggested that concerns about the preservation of web-based records may be as much about economic viability as they are about material infrastructure. The GeoCities collections discussed in this chapter stand as stark a reminder of the role that commercial service providers play in the longevity of personal records. The imminent deletion of GeoCities makes the collections of both Archive Team and the Internet Archive unique in this chapter; with little time and limited access, involving record creators in the web archiving process was less logistically feasible than it would be in either the Zine Web Archive or Mormon Missionary Collection.

The GeoCities collections are also unique in that they represent collecting at the level of the service, rather than the individual. Indeed, these collections are consistently noted for their valuable contribution to web history. The wholesale deletion of all extant GeoCities websites, Ian Milligan has attested, "would have meant a large gap in our collective understanding of the early web."[418] Framing these records in this way positions them as vital to our collective understanding of how the technologies and cultural practices of the web evolved. They are described in the aggregate, as a collection: "the GeoCities archive," suggestive of a corporate collection acquired from the company that maintained control of them. They are less frequently discussed in terms of the personal or individual. There is a tension at work here, between understanding these as personal websites and understanding them as GeoCities websites. The records created in these online systems nonetheless often fall within traditional archival notions of personal records: those created by individuals in the course of conducting their affairs outside of a professional setting. Indeed, the GeoCities collections are populated with personal records. Scott noted this in his talk at the

[418] Milligan, "Welcome to the Web," 140.

2011 Personal Digital Archiving conference, citing the personal value of a memorial website created by a mother who had lost her son.[419] Yet, without individual involvement, the collection is most frequently positioned as evidence of the early web.

The example of GeoCities' wholesale deletion and the subsequent efforts to preserve it exemplify the role that the socio-technical environments in which records are created, and the commercial services that provide them, have a profound impact on digital cultural heritage. Yahoo's decision to delete GeoCities when it ceased to be popular or profitable is by no means an isolated incident. It has, however, stood out as a case study in the impact of networked commercial infrastructure on personal digital archives. In the next chapter, the emergence of socially networked, cloud-based platforms, will further emphasize the role that commercial service providers play in the production and preservation of personal digital records.

[419] Scott, "The Splendiferous Story of Archive Team."

6.0 Personal Records in Networked Social Platforms

The collections discussed in this chapter contain personal records that were created and stored in networked social platforms, primarily in the late 2000s through the late 2010s. Like the records discussed in the previous chapter, these records were created online, stored remotely, and could be accessed from multiple devices. However, the records discussed in this chapter were created in social platforms, which differ from records found on the open web in ways that have further impacted archival practice. While many of these records can be accessed, to some degree, on the open web, their full content and context are accessible primarily to other users who have logged into their own accounts on the same platform. Within these semi-closed and closed systems, the role of the platform itself becomes increasingly integral to the ways in which records are created and collected. Collections discussed in this chapter include the #MeToo Collection at Harvard University; the Thomas S. Mullaney papers at Stanford University; the University of North Carolina at Chapel Hill Confederate Monument Protests Collected Tweets; the Ursula K. Le Guin papers at the University of Oregon; and the Twitter Archive at the Library of Congress. Particular attention will be dedicated to the Library of Congress's Twitter archive, which represents the first acquisition of a complete social media service's public archive directly from the platform itself. For complete descriptions of the scope and content of each of these collections, see Appendix A.3. The proceeding chapter is organized according to this dissertation's primary thematic lenses of Materiality; Custody and Control; and Privacy and Publicity.

6.1 Introduction

The records created in networked social platforms discussed in this chapter are similar to those explored in the previous chapter in that they are remotely stored and accessed, and yet, as Henniger and Scifleet have observed, they are part of a second generation of web-based infrastructure – which has often been referred to as Web 2.0 – that is "so distinct from Web 1.0 that they are hard to ascribe to recently established norms of digital acquisition, preservation, and dissemination."[420] As this chapter will demonstrate, while web archiving strategies can and are used to collect personal records created in networked social platforms, archivists have had to develop new practical strategies in order to collect the robust metadata that provides much of the context for these socially networked records. Before exploring the approaches taken to collect these records, it is necessary to describe what they are, and how they differ from the personal websites and blogs found on the open web.

In the mid-2000s and 2010s, web-based personal communication and documentation moved from the largely open web to partially open or closed spaces. These are referred to alternately in the literature as "social network sites," "social media services," or "social media platforms." These systems have been defined by danah m. boyd and Nicole B. Ellison, who use the term "social network sites," as:

> "web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system. The nature and nomenclature of these connections may vary from site to site."[421]

[420] Henniger and Scifleet, "How Are the New Documents of Social Networks Shaping Our Cultural Memory?" 279.
[421] danah m. boyd and Nicole B. Ellison, "Social Network Sites: Definition, History, and Scholarship," *Journal of Computer-Mediated Communication* 13, no. 1 (2008): 211.

As in the previous chapter, the availability of affordable, often monetarily free, storage has supported the creation and retention of a vast quantity of personal digital records, stored remotely "in the cloud." Social media platforms encourage record creation without requiring record deletion or other acts of personal appraisal. Amelia Abreu has observed that these "contemporary technologies have given us platforms that look a lot like personal archives, yet the archival functionality of platforms feels like an empty promise."[422] The availability of so much storage, as Jennifer Gabrys has argued, has enabled a pervasive belief that "we are able to store everything, but in that ambitious documentation, we at the same time inevitably include the decay and oblivion that, at one time, it was the task of the archive to guard against."[423] Instead, the affordances of seemingly infinite storage outweigh the "relevance of particular material as archive-worthy."[424] Personal records created within these closed, or partially closed, platforms, are vulnerable to the social, political, and economic conditions of those systems. Platforms may cease to be viable or relevant and disappear entirely, as with the example of GeoCities in the previous chapter. Conversely, there are issues that arise when platforms persist over time, serving as repositories of personal data, and in some instances, tools for surveillance. "Unless you feel a desire to engage with your past self," Abreu continued, "it's easy to leave these platforms alone and forget those versions of yourself ever existed, along with the troves of data associated with them."[425] Stored in the forgotten, cloud-based platforms, personal records function as big data. Indeed, Tung-Hui Hu has observed, "much of the cloud's data consists of our own data, the photographs and content

422 Amelia Abreu, "The Collection and the Cloud," *The New Inquiry*, March 9, 2015, https://thenewinquiry.com/the-collection-and-the-cloud/.
423 Jennifer Gabrys, *Digital Rubbish* (Ann Arbor: University of Michigan Press, 2013), 118.
424 Gabrys, Digital Rubbish, 119.
425 Abreu, "The Collection and the Cloud."

uploaded from our hard drives and mobile phones; in an era of user-generated content, the cloud is, most obviously, our cloud."[426] Recognizing both the cultural heritage value of the records stored in "our cloud," and the reality that long-term preservation and access are not the primary missions of platform providers, archivists have endeavored to collect and preserve these materials.

Concerns about the "digital Dark Ages" have persisted through the present day. Google's Vint Cerf famously revived this phrase in 2015 when he expressed his concerns about the long-term preservation of the "bits of information" created and stored in the cloud.[427] Cerf noted that current digital preservation options were insufficient to preserve the many records stored within this infrastructure. However, Cerf's warnings ignored the many efforts and initiatives undertaken by archivists to preserve these very records. As Bertram Lyons argued in a response to Cerf's remarks, archivists "are not and have not been absent from the digital preservation questions. We are, however, hidden in the public narrative."[428]

Indeed, archivists have been engaged in the collection and preservation of web-based records, including those created in networked social platforms, for as long as these records have been created. The preservation of personal digital records specifically has been a subject of interest for a number of archivists and archival scholars. As Henniger and Scifleet have argued, the personal records created in networked social platforms are "the authentic personal expression of individuals recording their everyday experience" and are thus well within the purview of archivists and other information professionals who have traditionally collected letters, diaries, and other

[426] Tung-Hui Hu, *A Prehistory of the Cloud* (Cambridge: The MIT Press, 2015), xvii.
[427] Lauren Maffeo, "Google's Vint Cerf on How to Prevent a Digital Dark Age," *The Guardian*, May 29, 2015, https://www.theguardian.com/media-network/2015/may/29/googles-vint-cerf-prevent-digital-dark-age.
[428] Bertram Lyons, "There Will Be No Digital Dark Age," *Issues & Advocacy*, May 11, 2016, https://issuesandadvocacy.wordpress.com/2016/05/11/there-will-be-no-digital-dark-age/.

personal records for the purposes of preserving cultural heritage.[429] Sinn and Syn have argued that archivists "cannot overlook this wealth of information about contemporary life."[430] As a growing portion of the contents of personal archives is created and stored in social media platforms, from Facebook to Twitter to Snapchat, a number of archival scholars have turned their attention to the consideration of how "platform functionalities" shape and affect user-generated content.[431] Acker and Brubaker recommend that archivists theorizing and collecting personal archives should "adopt a platform perspective that includes preserving the contextual integrity of networked data, confronting the shifts in the persistence of platforms, and clarifying archival expectations to provide access to personal collections created with social media platforms."[432] The past decade has seen a variety of approaches taken to the collection and preservation of personal records created in complex, networked systems.

The Library of Congress' 2010 acquisition of the Twitter archive, a gift from the social media platform, has been credited as the genesis of social media archiving.[433] The decision to preserve Twitter's public archive in its entirety has been described as a "tectonic shift" in the practices of memory institutions,.[434] Throughout the decade since the donation was announced, this collection has remained a touchstone for discussions about social media archives, generating debate about archival appraisal, processing, and ethics – and, in particular, the challenges of acquiring such a massive collection. In a 2016 article about the Library's struggle to manage the

[429] Henniger and Scifleet, "How Are the New Documents of Social Networks Shaping Our Cultural Memory?" 283.
[430] Donghee Sinn and Sue Yeon Syn, "Personal Documentation on a Social Network Site: Facebook, a Collection of Moments from Your Life?" *Archival Science* 14, no. 2 (2014): 119.
[431] Amelia Acker and Jed Brubaker, "Death, Memorialization, and Social Media: A Platform Perspective for Personal Archives," *Archivaria* 77 (2014): 1.
[432] Acker and Brubaker, "Death, Memorialization, and Social Media," 2.
[433] Seemantani Sharma, "'How Tweet it is!': Have Twitter Archives Been Left in the Dark?" *Journal of Law, Technology, and Policy* no. 1 (2019): 50.
[434] Henniger and Scifleet, "How Are the New Documents of Social Networks Shaping Our Cultural Memory?" 279.

collection, Andrew McGill wrote that "however dubious" the task of archiving Twitter may have seemed in 2010, "no one doubted the Library of Congress would get the work done. If Twitter could handle a few million tweets a day, surely the largest library in the world could, too."[435] As McGill and others have noted, however, this has proven much more difficult than anticipated. While the Library's Twitter archive has yet to be made accessible to researchers, other memory institutions have taken up the task of collecting personal records created and stored with Twitter.

Interest in social media archiving has grown, particularly in response to the use of hashtags to connect individual posts related to specific social and political movements. As a growing number of social justice movements have begun to be documented online, Brooke Morris-Chott has observed, "archivists are discovering the need to create collections of born-digital materials related to current events from both the fringe and mainstream."[436] In addition to collecting the official records and published accounts of current events, archivists have begun to collect records created by private individuals on social media platforms, and in particular Twitter, in order to document events from the perspective of the people who are experiencing them. Both the #MeToo Collection at Harvard University's Schlesinger Library and the Confederate Monument Protest Collected Tweets at the University of North Carolina at Chapel Hill are two examples of institutions collecting tweets affiliated with specific hashtag movements.

Efforts to collect the records created in networked social platforms note that the mission of these platforms is not to preserve and provide access to digital cultural heritage in the long term. A press release announcing the creation of the #MeToo collection at Schlesinger Library, for

[435] Andrew McGill, "Can Twitter Fit Inside the Library of Congress?" *The Atlantic*, August 4, 2016, https://www.theatlantic.com/technology/archive/2016/08/can-twitter-fit-inside-the-library-of-congress/494339/.
[436] Brooke Morris-Chott, "Archiving Controversial Digital Materials," Radcliffe Institute for Advanced Study News, August 8, 2019, https://www.radcliffe.harvard.edu/news/in-news/archiving-controversial-digital-materials.

example, cited the precarious state of social media records, stating that these "resources which are now pervasive in our collective consciousness and social media feeds, yet will prove acutely vulnerable in the long-term, as proprietary platforms, individual user-accounts, and the ever-changing landscape of the Web continually transform."[437] While these records have been "multiplying at an exponential rate on a daily basis for more than a year, it is acutely vulnerable over the long term. Quantity is not a guarantee of permanence."[438]

While the impetus to collect personal records created in networked social platforms, particularly where they relate to significant current events, has been clearly expressed, these records have also presented new challenges for archivists. Morris-Chott observed of the #MeToo Collection that "many of these materials contain controversial or sensitive content, and archivists face challenges related to collection scope, ethics, access, liability, contexts, appraisal, technology, and staff safety."[439] The records explored throughout this chapter may be understood to be sensitive because of the specific events or issues that they document; however, they must also be understood and treated with sensitivity simply because they are the personal records of private individuals. Community initiatives such as Documenting the Now have attempted to develop tools and educational materials in order to encourage thoughtful, ethical collection of social media content. The influence of their work can be observed in several of the collections examined in the proceeding chapter.[440]

[437] "Schlesinger Library Awarded Grant to Create Comprehensive Digital Media Archive of #metoo," Radcliffe Institute for Advanced Study Harvard University, June 25, 2018. https://www.radcliffe.harvard.edu/news/press-releases/schlesinger-library-awarded-grant-digital-archive-metoo.
[438] Amanda Strauss, "#MeToo: A Glimpse into the Digital Vault," Schlesinger Newsletter, Fall 2018, https://www.radcliffe.harvard.edu/news/schlesinger-newsletter/metoo-glimpse-digital-vault.
[439] Morris-Chott, "Archiving Controversial Digital Materials."
[440] Documenting the Now, https://www.docnow.io/.

6.2 Materiality

Much like the web-based records discussed in the previous chapter, the records created within networked social platforms are stored remotely, accessed remotely through personal computers, tablets, and smartphones. These records are often described as being stored "in the cloud," a euphemistic term that serves to obfuscate the reality that our records are actually, as Abreu has observed, "situated in a complex and obscure climate of for-profit data centers and server farms."[441] The cloud is media agnostic, rather than media-specific, as Hu has noted, a reality that makes it difficult to trace the specific technologies that support it.[442] The United States National Institute of Standards and Technology (NIST) has developed the following definition of cloud computing:

> "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."[443]

Within the infrastructure of the cloud, individuals are several degrees removed from their records by design; this system offers a sense of ease, and even, as in the previous chapter, a sense of immateriality. Indeed, even the label "the cloud," serves to obfuscate, as the cloud is actually many clouds: NIST details four possible cloud deployment models, designed to serve separate audiences and functions. These include private clouds, designed for use by a single organization;

[441] Abreu, "The Collection and the Cloud."
[442] Hu, *A Prehistory of the Cloud*, xix.
[443] Peter Mell and Timothy Grance, The NIST Definition of Cloud Computing, Special Publication 800-145, National Institute of Standards and Technology, https://www.profsandhu.com/cs6393_s13/nist-SP800-145.pdf, 1.

community clouds, designed for organizations that share specific concerns or functions; public clouds, designed for the general public and managed by business, academic, or government organizations; and hybrid clouds, which may be compositions of two or more of the aforementioned cloud infrastructures.[444] Far from being immaterial, the cloud simply refers to the deployment of these networked infrastructures, which may be governed by a variety of public and private institutions.

Within this infrastructure, the records created in networked social platforms do not appear to be fixed upon a stable medium. However, as Henniger and Scifleet have observed, "there is a materiality to the social networking service documents that is constituted by the architecture of the digital text" and "its supporting infrastructure."[445] Before proceeding to discussion of the forms in which these networked records have been acquired by archives, it is necessary to describe their characteristics as active records. Because so many of the collections discussed throughout this chapter contain records created in the social media platform Twitter, it is helpful to consider the structure of a tweet in order to illustrate the complexity of these records.

Figure 9 depicts the records created in Twitter, as they appeared to a user of the service in 2010. This figure illustrates the information that was visible to a user of the platform at the interface level.[446]

[444] Mell and Grance, The NIST Definition of Cloud Computing, 3.
[445] Henniger and Scifleet, 282.
[446] A representation of Twitter in 2010 has been chosen primarily because this was the year in which the Library of Congress acquired the Twitter archive; it also provides a historically appropriate comparison to Figure 10, which depicts the underlying infrastructure of a tweet, also in 2010. Bob Leggitt, "The History of Twitter in Profile Pages: 2006 to 2015," Twirpz, June 3, 2015, https://twirpz.wordpress.com/2015/06/03/the-history-of-twitter-in-profile-pages-2006-to-2015/.

Figure 9 Screenshot of a tweet in 2010

Figure 10, conversely, illustrates the underlying structure of a tweet, which is considerably more complex. Designed by former Twitter software engineer Raffi Krikorian in 2010, this diagram reveals the detailed metadata that accompanied a single tweet. Even this diagram belies the complexity of a Twitter record; in 2014 coverage of the Library of Congress's Twitter archive, it was reported that each tweet within the collection contained up to 150 unique metadata fields.[447] Given this infrastructure, archivists who wish to collect records created in social media platforms, such as Twitter, have had to make critical decisions about the specific properties that they wish to prioritize in their collections. Both web archiving strategies, such as those discussed in the previous chapter, and specific social media archiving strategies that rely on access to platform APIs have

447 Elizabeth Dwoskin, "In a Single Tweet, as Man Pieces of Metadata as There Are Characters," Wall Street Journal, June 6, 2014.

been used in the collections examined in this chapter. Each of these methods produces a distinct, unique archival record.

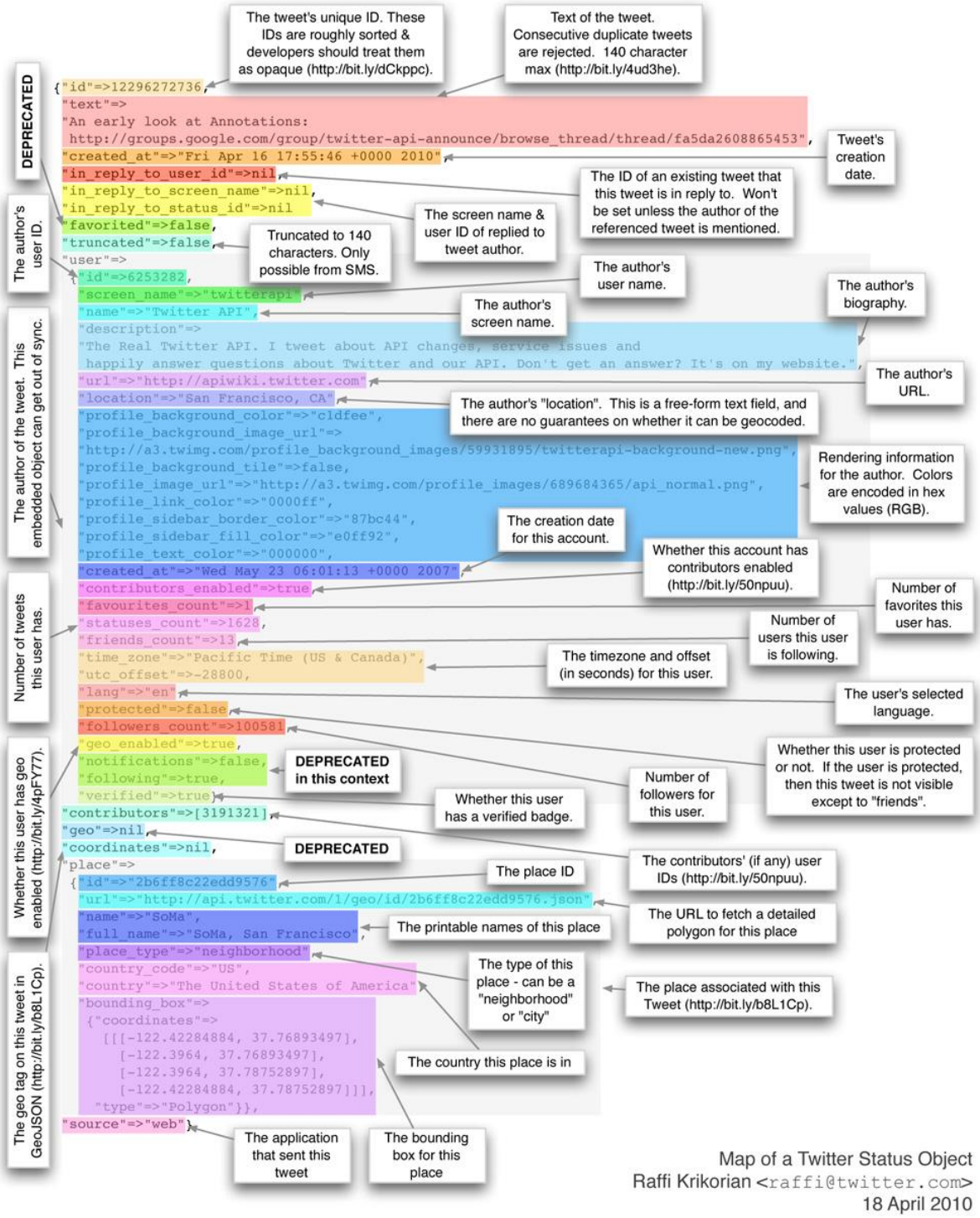

Figure 10 Map of a Twitter Status Object. Raffi Krikorian. 2011.

Web archiving methods, like those employed in collections discussed in the previous chapter, have been used to capture records created in social media platforms. Both the Ursula K. Le Guin digital materials at the University of Oregon and the Thomas S. Mullaney digital materials at Stanford, are examples of this approach. Each of these collections has been built with the use of the Archive-It web archiving service.[448] Using these methods, Justin Littman has explained, "the website delivers social media content as HTML (and CSS and JS and other) for rendering via browsers.[449] Web crawlers request content from web servers and "record" the files that the server returns to the web browser such that they can later be "played back to re-enact the experience of using the website at the time the capture occurred."[450] The resulting record, stored in WARC format within Archive-It's storage infrastructure, allows users to access the surface-level interface of the Twitter profile as it appeared at the time of capture.[451] Users may scroll through the tweets and retweets that appeared on Mullaney's profile, but may not click any of the links to view, for example, the other Twitter users who liked or retweeted any of his tweets. An attempt to click on the archived version of one of Mullaney's tweets renders the pop-up message: "Tweet does not exist." Likewise, users of Le Guin's archived Instagram profile may click on individual photos posted from her account, but cannot access the list of other Instagram users who liked those photos. They may not, additionally, access the list of the 8 Instagram users who Le Guin followed.

If one's objective is to capture more components of a tweet, including metadata, web crawling methods are not likely to be sufficient, as they are unable to capture the complete data

[448] Ursula K. Le Guin Website and Social Media, Archive-It, https://www.archive-it.org/collections/10533; Thomas S. Mullaney, Archive-It, https://archive-it.org/collections/10862.
[449] Justin Littman, "Web archiving and/or/vs Social Media API Archiving," Social Feed Manager, December 12, 2017. https://gwu-libraries.github.io/sfm-ui/posts/2017-12-13-web-social-media-archiving.
[450] Littman, "Web archiving and/or/vs Social Media API Archiving."
[451] Archived version of Thomas S. Mullaney's Twitter profile, https://wayback.archive-it.org/10862/20181203234208/https://twitter.com/tsmullaney.

and metadata that comprise the record.[452] In order to collect records that include these structural components, archivists have begun to use tools and techniques that have been designed specifically to capture the rich contextual information contained within the metadata of social media records. Though these approaches share the objective of capturing content born on the web for long-term stewardship, web archiving and social media API archiving produce records that are markedly different, both structurally and in the information they convey. Social media archiving methods make use of the application programming interface (API) of a social media service. Archivists can create requests and the API returns data that matches the query, often in the form of JSON datasets.[453] This method returns structured data only, as opposed to web archiving methods which typically capture the "look and feel" of web-based records at the level of the interface. The #MeToo Collection and the Confederate Monument Protest Collected Tweets both consist of tweet data associated with specific hashtags that has been collected through the use of Twitter's API.

Littman has suggested that ideally, "web archives and social media API archives should be thought of as complementary," because they each capture distinct and unique components of active social media records.[454] However, if collecting multiple forms of each record is not feasible, archivists must choose the strategy that will capture the properties of records that are most significant to them. Significant properties are "those properties of digital objects that affect their quality, usability, rendering, and behavior," Margaret Hedstrom and Christopher Lee have explained.[455] They add that while, in an ideal world, "free from technical and economic

[452] Justin Littman, Daniel Chudnov, Daniel Kerchner, Christie Peterson, Yecheng Tan, Rachel Trent, Rajat Vij, & Laura Wrubel, "API-Based Social Media Archiving as a Form of Web Archiving," *International Journal on Digital Libraries* 19 (2018): 21-38.
[453] Littman, "Web archiving and/or/vs Social Media API Archiving."
[454] Littman, "Web archiving and/or/vs Social Media API Archiving."
[455] Margaret Hedstrom and Christopher A. Lee, "Significant Properties of Digital Objects: Definitions, Applications, Implications," *Proceedings of the DLM Forum* 2002, https://ils.unc.edu/callee/sigprops_dlm2002.pdf, 218.

constraints," archivists would be able to preserve their collections without sacrificing any of these properties, such constraints do exist, and archivists must determine which properties will be most meaningful to their users.[456] Social media archives created with the use of APIs, conversely, have proven more useful to researchers seeking large scale datasets related to specific phenomena, particularly if they wish to apply computational methods in order to interpret them. In some instances, as in the Confederate Monument Collected Tweets, only tweet identifiers have been made available, rather than the complete data and metadata of the original tweets; users of the collection can access the complete tweet data via the Twitter API.[457] As will be discussed further in subsequent sections, the terms of social media APIs impact not only the structure of the records that are collected, but the ways in which archivists can acquire those records, and the terms under which they may be made accessible to researchers.

Of the collections discussed in this chapter, only the Twitter archive at the Library of Congress contains all public-facing data created within an entire platform, including the metadata for those records. The volume of this data has proven increasingly challenging to the archivists responsible for processing it. In 2010, users of the social media platform posted a total of approximately 55 million tweets per day.[458] In February 2012, the Library received its first bulk data transfer from Twitter, consisting of all public tweets from 2006 through 2010 in three compressed files that totaled 2.3 terabytes; uncompressed, the files reached 20 terabytes, or approximately 21 billion tweets, each with over 50 accompanying metadata fields.[459] An update

[456] Hedstrom and Lee, "Significant Properties of Digital Objects," 218.
[457] "Restrictions to Access," University of North Carolina at Chapel Hill Confederate Monument Protests Collected Tweets, 2017-2020, https://finding-aids.lib.unc.edu/40486/#d1e88.
[458] Andrew McGill, "Can Twitter Fit Inside the Library of Congress?" *The Atlantic*.
[459] Michael Zimmer, "The Twitter Archive at the Library of Congress: Challenges for Information Practice and Information Policy," *First Monday* 20, no. 7 (2015), https://firstmonday.org/article/view/5619/4653.

on the project noted that by February 2011, the Library was acquiring the approximately 140 million tweets created each day, and that by October 2012, that number had risen to "nearly half a billion tweets each day."[460] By 2013, the Library had acquired over 170 billion tweets, or 133 terabytes of data, and the collection was growing rapidly.[461] The increased production in tweets has often been cited as a cause for the project's slow development. Each tweet in the Library's collection is stored as a JSON file.[462]

Early in the Twitter archiving project, the web archiving team reported "budgetary pressures, limited access to server disk space, [competition] with other big data projects, and new infrastructure for new capabilities" as challenges associated with collecting social media records at large scale. By the time the goal of collecting all public tweets was abandoned in 2017, the Library cited the changing nature of Twitter, and particularly its increasing focus on multimedia content, which was excluded from the text-only collection, as a reason for its shift in scope.[463] Where tweets in 2010 had consisted primarily of text and metadata, in the years to come they would expand to include "embedded photos, then video, and then live video," resulting in additional metadata that "weighed down the Library of Congress's daily downloads, and forced staff to consider building an archival system that would change as often as Twitter did."[464] Changes in the material infrastructure of Twitter, including the increase of posting photographs and videos

[460] Gayle Osterberg, "Update on the Twitter Archive at the Library of Congress," Library of Congress Blog, January 4, 2013, https://blogs.loc.gov/loc/2013/01/update-on-the-twitter-archive-at-the-library-of-congress/.
[461] Osterberg, "Update on the Twitter Archive at the Library of Congress," 2013.
[462] Audrey Watters, "How the Library of Congress is Building the Twitter Archive," *Radar*, June 2, 2011, http://radar.oreilly.com/2011/06/library-of-congress-twitter-archive.html.
[463] "Update on the Twitter Archive at the Library of Congress," White Paper, December 2017, https://blogs.loc.gov/loc/files/2017/12/2017dec_twitter_white-paper.pdf.
[464] McGill, "Can Twitter Fit Inside the Library of Congress?"

in addition to text, and posting embedded media that is hosted outside of Twitter, were cited as further justification for re-envisioning the scope of the project.[465]

The socio-technical environments in which personal records are created have a clear impact on the information provided in these records, particularly with regard to the contextual information provided by record metadata. In turn, web archiving and social media archiving strategies shape the record that is preserved within archival institutions. As the following section will discuss, the terms under which social media platforms permit the collection of records created within them will also have a profound impact on collecting practices and the means through which archivists may make these collections accessible to their users.

6.3 Custody and Control

"Today, most of our personal digital memory is not under our control," Smith Rumsey observed in her recent work *When We Are No More: How Digital Memory is Shaping Our Future*. "Whether it is personal data on a commercially owned social media site, e-mails that we send through a commercial service provider, our shopping behaviors, our music libraries, our photo streams," are tied up in a remote corporate infrastructure that extends beyond individual control.[466] While this is an issue that undoubtedly concerns individual record creators, it is also one that concerns archivists. Smith Rumsey posed the question, "who has the right to preserve digital content on behalf of the public – present and future? Do we own our personal data – biomedical,

[465] "Update on the Twitter Archive at the Library of Congress," White Paper, December 2017,
[466] Smith Rumsey, *When We Are No More*, 144.

demographic, political – and can we control its use?"[467] As a growing number of personal records have been created in commercial platforms, archivists have been compelled to consider the parts they will play in determining the ownership, control, and long-term care of these materials. The collections examined in this chapter exemplify three approaches that archivists have taken to acquiring custody of personal records that have been created in networked social platforms. Given the complex sociotechnical infrastructure within which these records have been created and stored, archivists have employed methods of assuming custody or control of personal records that are both new and familiar. Some draw on the same web archiving strategies used to collect records from the open web, as in the previous chapter, while others rely on social media archiving strategies that utilize platform APIs, as discussed in the previous section. The Twitter archive at the Library of Congress stands as a lone example of a formal transfer of custody between a service provider and a memory institution. While the Library's Twitter archive remains something of an anomaly in the field, it has been a touchstone for discussions about the ownership and control of personal records that are created within commercial infrastructure. The influence of the Terms of Service established by commercial service providers thus emerges as an area of concern for archivists.

Table 7 Methods of acquiring socially-networked personal digital records

Collection	Collecting Institution	Acquisition Method	Acquisition Source
#MeToo Collection	Harvard University	Web Crawling; Crowdsourcing	Open web; User submissions
Thomas S. Mullaney Papers	Stanford University	Legal Transfer of Custody; Web Crawling;	Creator; Open web
Twitter Archive	Library of Congress	Transfer of Custody	Service Provider

[467] Smith Rumsey, *When We Are No More*, 144.

Ursula K. Le Guin Papers	University of Oregon	Legal Transfer of Custody; Web Crawling	Creator; Open web
Confederate Monument Protests Collected Tweets	University of North Carolina at Chapel Hill	Twarc	Service API

The legal terms under which records are collected likewise shape collecting practices in social media archives. Terms of service, alternately referred to as terms of use or terms and conditions, are the "rules a person or organization must observe in order to use a service."[468] These documents establish the terms of ownership and control of records created within the platform, which are subject to change at regular or irregular intervals throughout the lifespan of a service provider. Twitter, for example, reserves the right to change their terms at any time, and notification to users is provided, or not provided, at Twitter's discretion.[469] They are also, as it has been frequently observed, lengthy documents, typically written in complex legal and technical language; as a result, many users have only a tentative understanding of the terms to which they agree when creating their user accounts.[470] As Hugo Roy has argued, "what's problematic is that although terms of service are kinds of non-negotiated contracts, they're still a two way relationship. There are meaningful things in there, about [the user's] rights and obligations as well, to which you presumably consent."[471] The lack of transparency and legibility in terms of service are of direct consequence to archivists who collect personal records created within social media platforms. The majority of collections discussed within this chapter have stated their adherence to

[468] "Terms of Service," PC Magazine Encyclopedia, https://www.pcmag.com/encyclopedia/term/terms-of-service.
[469] Hugo Roy, "'What is Yours Stays Yours': Oh, Really?" Terms of Service; Didn't Read Blog, June 17, 2013, https://tosdr.org/blog/tosdr-in-action-owned.html.
[470] Hugo Roy, "ToS;DR in Action: 'I Have Read and Agree to the Terms," Terms of Service; Didn't Read, June 10, 2013, https://tosdr.org/blog/tosdr-in-action-i-have-read-and-agree.html.
[471] Roy, "ToS;DR in Action: 'I Have Read and Agree to the Terms."

the terms of service of the platforms from which they are collecting records, an approach that specifies adherence to the legal conditions under which acquisition is permitted. Twitter's donation to the Library raised, for many Twitter users, the question of "who owned the data and had the right to decide what to do with it. Many people were unpleasantly surprised to realize that they did not control their Twitter streams," despite Twitter's assurances that users were the owners of all content that they created within the platform.[472] A closer look at Twitter's terms of service reveals that ownership and control – as in many archival agreements – are in fact two separate issues.

APIs play a significant role in the acquisition of personal records created in networked social platforms, and particularly the ways in which physical custody of records is established. Additionally, each platform's terms of service play a role in shaping collecting practices.

The Library of Congress, Harvard University, and University of North Carolina at Chapel Hill have each addressed the ways in which they adhere to Twitter's terms of service in their own collecting practices. Of the collections discussed in this chapter, only the Library of Congress's Twitter archive was acquired through a formal agreement between service provider and collecting institution. The two-page deed of gift – the "formal and legal agreement between the donor and the repository that transfers ownership of and legal rights to the donated materials" – establishing the terms of this agreement is remarkably brief given the scope and complexity of the collection.[473] Indeed, archival scholar and Documenting the Now Project Director Bergis Jules has called this deed of gift "almost laughable in its simplicity. Jules also noted, however, that as the first agreement of its kind, it could nonetheless serve as an example to future collections that included

[472] Abby Smith Rumsey, *When We Are No More*, 139.
[473] "A Guide to Deeds of Gift," Society of American Archivists, https://www2.archivists.org/publications/brochures/deeds-of-gift#.V0kj984sj6U. The gift agreement for the Twitter archive at the Library of Congress can be found at https://blogs.loc.gov/loc/files/2010/04/LOC-Twitter.pdf.

social media data.[474] This agreement largely defers to the terms established by Twitter's Terms of Service as they existed at the time of the agreement; as of April 2010, this meant Version 3 of the Terms of Service.[475] The gift agreement granted "an irrevocable nonexclusive license to the Library for such rights as the Donor [Twitter] has the right to transfer or license under the Twitter Terms of Service at the time of the gift or before."[476]

In Version 3 of the Terms of Service, Twitter states that users retain their rights to any content that they "submit, post, or display" in the platform. However, they continue:

> "By submitting, posting, or displaying Content on or through the Services, you grant us a world-wide, non-exclusive, royalty-free license (with the right to sublicense) to use, copy, reproduce, process, adapt, modify, publish, transmit, display and distribute such Content in any and all media or distribution methods (now known or later developed)."

In other words, as summarized in the Terms, "This license is you authorizing us to make your Tweets available to the rest of the world and to let others do the same. But what's yours is yours – you own your content."[477] Users may own their ideas, the Terms suggest, but not the manifestations of them stored within Twitter's data centers. While content may be owned by the individual creator, it is controlled by Twitter. This includes the provision of access, by way of its API, to archives and other collecting institutions.

While other institutions have not entirely replicated the Library's Twitter archive by collecting the complete public archive of a social media service, it has become increasingly common to find records from social platforms in archival collections. The Twitter archive has, in

[474] Bergis Jules, "Some Thoughts on Ethics and DocNow," DocNow News, June 3, 2016, https://news.docnow.io/some-thoughts-on-ethics-and-docnow-d19cfec427f2.
[475] "Version 3," Twitter Terms of Service, https://twitter.com/en/tos/previous/version_3.
[476] Twitter archive gift agreement, https://blogs.loc.gov/loc/files/2010/04/LOC-Twitter.pdf.
[477] https://twitter.com/en/tos/previous/version_3.

many respects, served as a precedent for approaches to social media archiving. Indeed, four of the five collections explored in this chapter features personal records created in Twitter.

The #MeToo collection at Schlesinger Library, for example, This collection also, however, invites individuals to submit their own electronic materials to be considered for inclusion, allowing for a transfer of custody between individual creators and archives as well.[478] The collection team is clear about what a donation to the collection means, stating on their submission form that they will request a "non-exclusive right to: preserve, reproduce, display, distribute, and make the material you submit publicly available on the web."[479] Non-exclusive rights mean that the Schlesinger is not asking for the only copy of the submitted records; contributors will be able to "grant, assign, and retain" their original rights to the records."[480]

The finding aid for the Confederate Monument Protest Tweets at UNC Chapel Hill explains that tweets in its collection were acquired using the Twitter API and the twarc python package, a resource supported by Documenting the Now. Twitter's API allows users to search tweets from the previous seven days, a limitation that shapes the scope of collections that rely on API archiving tools.[481] The finding aid notes that it has made only tweet identifiers, rather than complete tweet data, available because Twitter's Terms of Service prohibit them from making the complete data collect available to the public.[482] Users of the collection can use twarc to "hydrate" the tweet identifiers, gaining access to the complete tweet data.

[478] "Contribute," #MeToo Collection, https://www.schlesinger-metooproject-radcliffe.org/contribute.
[479] "Contribute to the #metoo Project at Schlesinger Library," #MeToo Collection, https://harvard.az1.qualtrics.com/jfe/form/SV_73eGeqtEoHPSsh7.
[480] "Contribute to the #metoo Project at Schlesinger Library," #MeToo Collection.
[481] "Acquisitions Information," University of North Carolina at Chapel Hill Confederate Monument Protests Collected Tweets, https://finding-aids.lib.unc.edu/40486/.
[482] "Restrictions to Access," University of North Carolina at Chapel Hill Confederate Monument Protests Collected Tweets.

Abiding by a commercial platform's terms of service is a way of working in partnership with the platform; but if archivists wish to build a relationship with the individuals who create records within these platforms, additional measures may be worth considering, given the confusion that surrounds these legal terms. Jules has asked whether archivists should "continue to treat social media archives like they have no owners when even Twitter acknowledges that users retain the rights to content in their tweets?"[483]

6.4 Privacy and Publicity

The question of privacy is complicated within social network sites, many of which offer tiered, customizable privacy settings for users, ranging from the very public to the very private. Even records that are set to a fully public status by their creators may be visible only to other users of the platform, accessible only when they are logged into their own user accounts. Legal scholar Chloé S. Georas has suggested that while "almost every archival effort must contend with the legal hurdle of copyright… the archiving of social networks must also address how to handle the potentially sensitive nature of materials that are considered 'private' from the perspective of the social and legal constructions of privacy."[484] While the focus of this dissertation is on social and archival understandings of privacy, rather than on its legal meanings, it is important to acknowledge the ways in which archives and other memory institutions have deferred primarily to privacy and copyright restrictions as they have been established by the terms of service and privacy

[483] Jules, "Some Thoughts on Ethics and DocNow."
[484] Chloé S. Georas, "Networked Memory Project: A Policy Thought Experiment for the Archiving of Social Networks by the Library of Congress," *Laws* 3 (2014): 470.

policies of social network sites. Archivists are often required to agree to a platform's Terms of Service in order to collect data, particularly when doing so through their APIs.[485] This section will consider the implications of these decisions, as well as efforts made to take a more social, personal approach to protecting privacy in collections that contain the social media records of individuals.

From the date of its initial acquisition of the Twitter archive, representatives from both Twitter and the Library of Congress emphasized the fact that the archived collection would include only those tweets that had been posted publicly to the platform; neither tweets from accounts that were set to private status nor tweets that had been deleted by their creators would be among those tweets that were collected and preserved.[486] In an announcement about the donation, Twitter Co-founder Biz Stone noted that "only a tiny percentage of accounts are protected," suggesting that most tweets were "created with the intent that they will become publicly available."[487] The acquisition fell within Twitter's Terms of Service at that time, which gave Twitter the right to make user-posted content "available to other companies, organizations, or individuals who partner with Twitter for syndication, distribution, or publication of such Content on other media services," with no requirement to notify or compensate users.[488] The version of Twitter's Privacy Policy in effect in April 2010 stated that public tweets were "searchable by many search engines and… immediately delivered via SMS and our APIs to a wide range of users and services."[489] Users, the policy suggested, "should be careful about all information that will be made public by Twitter, not

[485] Ed Summers, "A Twitter ToS Deep Dive with Justin Littman," DocNow Blog, March 7, 2019, https://news.docnow.io/a-twitter-tos-deep-dive-with-justin-littman-728feb54c0fd.
[486] Matt Raymond, "The Library and Twitter: An FAQ," Library and Congress Blog, April 28, 2010. http://blogs.loc.gov/loc/2010/04/the-library-and-twitter-an-faq/.
[487] Biz Stone, "Tweet Preservation," Twitter Blog, April 14, 2010, https://blog.twitter.com/official/en_us/a/2010/tweet-preservation.html.
[488] "Version 3," Previous Twitter Terms of Service.
[489] "Version 2," Previous Twitter Privacy Policy," https://twitter.com/en/privacy/previous/version_2.

just your Tweets."[490] That information might be details taken from a user's profile, and metadata including location data, tweets marked as "favorites" or retweeted, accounts following and followed by the user, and "many other bits of information."[491] Indeed, the Privacy Policy allows for the acquisition of the Twitter archive by the Library of Congress, and yet many users were surprised and dismayed to learn that they had little recourse when it came to their inclusion in the archives. The 107 user comments on the Library's announcement convey a range of responses, with several posters noting their concerns about privacy. One commenter wrote, "so with no warning, every public tweet we've ever published is saved for all time? What the hell. That's awful. We should have been warned about this. Now people will be able to look up our tweets for the rest of our lives and there's no way we can have them removed. Even if our tweets aren't bad or anything, this is hugely inappropriate."[492] Other commenters suggested that it was incumbent upon users to have read the Terms of Service and Privacy Policy before using the service: "I think the people who are concerned about their Tweets being available for all the world to see and that Library of Congress is now archiving these tweets should have read the Terms & conditions on Twitter - it seems it's all included there."[493] And a number of commenters argued that the expectation of privacy in any online space was a mistake on the part of users. For example, one commenter wrote that, "Clearly some of you haven't been paying attention. Do you really think you own any rights to information you post on the web? Privacy doesn't exist when you willingly post something to a public site."[494]

[490] "Version 2," Previous Twitter Privacy Policy."
[491] "Version 2," Previous Twitter Privacy Policy."
[492] Comment 33761, "How Tweet It Is!" Library of Congress Blog, April 14, 2010, https://blogs.loc.gov/loc/2010/04/how-tweet-it-is-library-acquires-entire-twitter-archive/#comment-33761.
[493] Comment 35502, "How Tweet It Is!" Library of Congress Blog, April 14, 2010, https://blogs.loc.gov/loc/2010/04/how-tweet-it-is-library-acquires-entire-twitter-archive/#comment-35502.
494 Comment 3913, "How Tweet It Is!" Library of Congress Blog, April 14, 2010, https://blogs.loc.gov/loc/2010/04/how-tweet-it-is-library-acquires-entire-twitter-archive/#comment-39313.

The perspective that the records created in social media platforms are "already public" and can thus be collected, preserved, and distributed without the individual consent of users has not been limited to the comments section. Indeed, it has been at the center of much debate in archival scholarship. Michael Zimmer, who has conducted extensive ongoing research on social media archives and the Library's Twitter archive, has argued that although a Twitter user may have set their account to public, a setting which allows others to search for, view, or follow their activity on the platform, "there is a reasonable expectation that one's tweet stream will be 'practically obscure,' within the thousands (if not millions) of tweets similarly publicly viewable."[495] The user may imagine or expect that only other individuals within their own immediate or extended networks will view their records. While users of Twitter have consented to making their tweets public within the live platform, this does not, Zimmer has argued, imply an informed consent to have their tweets "systematically followed, harvested, archived, and mined by researchers (no matter the positive intent of such research)."[496] Though records created in Twitter may be set to a "public" status, taking this status as consent to long-term preservation within an archival repository suggests a binary relationship between public and private, a perspective that ignores the kinds of "contextual norms" advanced by Nissenbaum.[497]

In the decade since the Library's acquisition of the Twitter archive, a growing number of memory institutions have acquired and actively collected personal records created in social media platforms, and Twitter in particular. In many respects, particularly with regard to privacy, the Twitter archive has served as a case study for archivists. Both the #MeToo collection and the

[495] Michael Zimmer, "Is it Ethical to Harvest Public Twitter Accounts without Consent?" Michael Zimmer, February 12, 2010. https://www.michaelzimmer.org/2010/02/12/is-it-ethical-to-harvest-public-twitter-accounts-without-consent/.
[496] Zimmer, "Is it Ethical to Harvest Public Twitter Accounts without Consent?"
[497] Michael Zimmer, "The Twitter Archive at the Library of Congress."

Confederate Monument tweet collection have stated their adherence to Twitter's Terms of Service. Both institutions responsible for these collections have also made well-documented efforts to approach their work ethically, and with a degree of care that extends beyond the official terms established by Twitter's policies.

In a presentation made at the 2019 Society of American Archivists annual conference, the #MeToo project team addressed the ways in which they have enacted their ethical codes into their work. Samantha Abrams advised that archivists collecting records created in social media platforms and other online spaces, including "who the work empowers, the speed and timing in which the materials are collected, who will be affected by the collection, where the content is best suited, and how access to the collection is provided."[498] These are questions that push beyond a binary understanding of private and public, and instead ask archivists to consider the multitude of potential effects of taking personal records from one context – i.e. the online environment in which they were created – and into another – the long-term storage repositories of cultural heritage institutions.

In its Ethics Statement, the #MeToo collection noted its adherence to the Terms of Service established by the commercial platforms in which records are created, and specifically, their commitment to respecting those terms with regard to the distribution of any data collected under the purview of the project.[499] Additionally, the #MeToo project team has cited the influence of the ethical codes established by the American Library Association and the Society of American Archivists in the development of their own policies. Among their stated ethical values is a commitment to transparency, a value that is arguably enacted through their provision of a publicly

[498] Morris-Chott, "Archiving Controversial Digital Materials."
[499] "Ethics Statement," #MeToo Collection, https://www.schlesinger-metooproject-radcliffe.org/ethics-statement.

accessible Ethics Statement, documentation of the research that has informed their ethical codes, and even a Zotero library populated with readings on ethics in web and social media archiving.[500] Indeed, their collection website, conference presentations, and publications have been used as a space to communicate their "goals, practices, and collections, including how archived materials can be accessed and under what conditions."[501] Of the records included within the collection, only the materials collected with Archive-It are currently accessible to the public online.[502] According to the collection website, Twitter data will be made available in 2020; it is unclear at the present through what means collected tweets will be made available to researchers. For access to materials other than those in the Archive-It collection, the #MeToo project team maintains a dedicated email address and asks potential users to "contact the project team with questions about this work."[503] Within the present system, questions and concerns, including those related to privacy, can be addressed on a case-by-case basis.

In December 2017, the Library announced that it would no longer collect all public tweets, opting instead for a selective appraisal strategy more closely aligned with its collecting policies for other web-based materials. In an announcement regarding the change, Osterberg explains that "the social media landscape has changed significantly, with new platforms, an explosion in use, terms of service and functionality shifting frequently and lessons learned about privacy and other concerns."[504] However, Osterberg does not explicitly describe the "lessons learned about privacy," nor are they addressed in the accompanying white paper. The white paper emphasizes that it has

[500] "Ethics Statement – Bibliography," #MeToo Collection, https://www.schlesinger-metooproject-radcliffe.org/ethics-bibliography.
[501] "Ethics Statement," #MeToo Collection.
[502] #MeToo Web Archives Collection, Archive-It, https://archive-it.org/collections/10866.
[503] "Ethics Statement," #MeToo Collection.
[504] Gayle Osterberg, "Update on the Twitter Archive at the Library of Congress," *Library of Congress Blog*, December 26, 2017, https://blogs.loc.gov/loc/2017/12/update-on-the-twitter-archive-at-the-library-of-congress-2/?loclr=blogloc.

never been the Library's mission to "collect comprehensively. Given the unknown direction of social media when the gift was first planned, the Library made an exception for public tweets. With social media now established, the Library is bringing its collecting practice more in line with its collection policies."[505] Collection policies for web archives at the Library of Congress emphasize a more selective appraisal policy, a marked shift away from the platform-level collecting that characterized the first seven years of the Twitter archive.[506] Importantly, these policies also entail an opt-in approach to collecting.

6.5 Conclusion

The collections studied in this chapter demonstrated that social media records, including those created in Twitter and Instagram, may also be collected through the web archiving methods detailed in Chapter Five. Web archiving and social media API archiving employ different methods in order to produce markedly different records. Archivists may choose to use one approach over the other because of the specific components captured by each. For example, the visual interface of a social media platform and the aggregated data collected through its API may potentially expose different data and metadata elements.[507] Each of the collections examined in this chapter contains personal records that have been created within networked social platforms, and most frequently, the social media platform Twitter.

[505] Osterberg, "Update on the Twitter Archive at the Library of Congress."
[506] "Collections Policy Statements and Overviews," Library of Congress. https://www.loc.gov/acq/devpol/.
[507] Littman, "Web archiving and/or/vs Social Media API Archiving."

Indeed, memory institutions appear to have recognized and accepted Twitter's role in supporting the production of digital cultural heritage. Taken on an individual basis, an individual user's tweets provide "evidence of me;" taken collectively, they constitute "evidence of us," in McKemmish's formulation. Indeed, Twitter can be understood as a space in which "our lives are individually and collectively witnessed and memorialized."[508] In 2011, McKemmish revisited of her seminal work, suggesting that "as we move further into the virtual world of technologies, new ways of negotiating the relationships between records and other forms of recorded information are emerging."[509] The result is that personal records may take multiple forms, perform multiple functions, and have "simultaneous or parallel provenances."[510] The collections examined in this chapter are exemplary of the recordkeeping paradigm described by McKemmish. While many of the tweets contained within these collections are personal records – documentation of the activities and experiences of individuals in the course of their lives – they also provide documentation that contributes to larger historical narratives, and, as in the Library of Congress's Twitter archive, a meta-narrative about the evolution of the platform itself.

The Twitter archive at the Library of Congress is unique among the collections studied in the preceding pages. Indeed, at the time of this writing it is the only known instance of a memory institution acquiring the complete public archive of a social media platform. It is similar to the GeoCities collections discussed in the previous chapter in that it constitutes an archival effort focused on the records of a specific platform; however, the terms of its acquisition distinguish it even from other platform-level collecting efforts. Further, these collections are similar in that they

[508] Sue McKemmish, "Evidence of Me… In a Digital World," in *I, Digital: Personal Collections in the Digital Era*, edited by Christopher A. Lee. (Chicago: Society of American Archivists, 2011): 117.
[509] McKemmish, "Evidence of Me… In a Digital World," 126.
[510] McKemmish, "Evidence of Me… In a Digital World," 126.

both contain multiple narratives, multiple histories. In a 2013 update on the Library's Twitter archive, Osterberg reiterated the Library's reasons for acquiring the collection, emphasizing that, "an element of our mission at the Library of Congress is to collect the story of American and to acquire collections that will have research value. So, when the Library had the opportunity to acquire an archive from the popular social media service Twitter, we decided that this was a collection that should be here."[511] Though the Twitter archive was mission-appropriate in this sense, it deviated significantly from existing collecting, and web archiving practices. Not only was the quantity of data massive, and structurally unlike the other born-digital collections held by the Library, but the acquisition of the aggregated records directly from Twitter was a step away from the permission-based "opt-in" approach taken with other web-based collections.

In late 2017, the Library announced that it would reconfigure its Twitter archive to more closely align it with other web archiving initiatives currently in place, and cease collecting all public tweets in favor of a more selective appraisal strategy.[512] At this time, the collection remains closed, with no indication if or when it will be made accessible to researchers. Critics of the project have suggested that under better leadership, the Library, "might have assessed the problems created by the initial Twitter acquisition and gone on to develop the tools and policy needed to create a useful collection."[513] However, examining other, smaller collections that include or consist entirely of Twitter data suggests that the project's flaws are not solely the product of the Library's mismanagement. Instead, this research has made clear the complexities of collecting personal records that are deeply embedded within commercially controlled, networked platforms.

511 Osterberg, "Update on the Twitter Archive at the Library of Congress," 2013.
512 Osterberg, "Update on the Twitter Archive at the Library of Congress," 2017.
513 Scott McLemee, "It Seemed Like a Good Idea at the Time." *Inside Higher Ed*. January 5, 2018, https://www.insidehighered.com/views/2018/01/05/essay-library-congresss-twitter-archive.

Indeed, archivists appear to be working within a complex network of parameters. If records take multiple forms and have multiple provenances, they are subject to multiple constraints. Archivists must comply with the Terms of Service established by platforms, as well as those of the platform APIs when data is collected through social media archiving methods. Additionally, archivists have attempted to develop ethical approaches to collecting personal records from networked platforms; the approaches taken in the #MeToo collection exemplify a thoughtful, intentional effort at treating sensitive data ethically. At the same time, social media archiving through APIs cannot provide an opt-in regime for individuals; within the parameters established by the API, these methods ultimately serve to protect the interests of the platform first.

Though this sample is too small to be representative, another pattern emerged in the course of this study. The Le Guin and Mullaney collections, the only collections in this chapter which were established on the basis of a single creatorship, both employed web archiving methods, and specifically the use of Archive-It, to collect records created in social media platforms. Social media archiving with APIs, conversely, has been used to collect records associated with specific hashtags, as in the #MeToo Collection and the Confederate Monument Protests Collected Tweets. Though not conclusive, this observation raises questions about the affordances and limitations of different tools for capturing networked social records, and the priorities and values embedded within them.

That these personal records are situated within a commercial recordkeeping infrastructure complicates archival approaches to collecting them, particularly with regard to privacy. While there are many who have argued, as was argued of records found on the open web, that tweets and other social media records are "already public," this argument is less convincing in the context of social platforms. Indeed, because these sites are predicated on connections between a user and their existing networks, as boyd and Ellison have argued, they occupy a space somewhere between

the extreme ends of publicity and privacy.[514] Further, as Zimmer has suggested, many users of social media services may expect that, within the confines of the platform, their records will be "practically obscure," and thus protected, even if they are technically set to "public" status.[515]

As this chapter concludes, it must be acknowledged that the records included in the collections examined in these pages represent only a small, and very specific, sample of the personal records being created in networked social platforms today. Archivists have not yet had to comprehensively address the implications of social media trends toward the ephemeral. Records created within Snapchat and Instagram Stories, for example, may be inaccessible to archives; yet it is often unclear how long these records are maintained as personal data with a platform's data centers. As time-based, "self-destructing" records remain popular, new questions arise for archivists. These seemingly ephemeral records prompt deeper consideration about the retention of personal records by commercial service providers, and the role that these providers will continue to play in the preservation of digital cultural heritage. Finally, and perhaps most interestingly, the tendency of individuals to create records that are designed to be forgotten raises further questions about the potential for personal records to resist archival notions of recordkeeping.

[514] boyd and Nicole B. Ellison, "Social Network Sites," 212.
[515] Zimmer, "Is it Ethical to Harvest Public Twitter Accounts without Consent?"

7.0 Conclusion

7.1 Introduction

In "The Perfect and the Possible: Becoming a Digital Archivist," Richard Pearce-Moses predicted that, as more and more digital records came to be found within archives, "*what* archivists do will not change, but *how* we do it may in fact change very dramatically."[516] The collections explored throughout this study support that notion, in a variety of ways.

This study considered archival treatments of personal records throughout the history of the profession in an effort to contextualize the examination of archival collections that contain personal digital records created and stored within three socio-technical environments: on personal computers; on the World Wide Web; and within social media platforms. The collections analyzed throughout the previous chapters suggest that archival collecting practices are shaped, in part, by the socio-technical infrastructures that support these records prior to their transfer to archival custody. They also demonstrate that within the constraints imposed by those infrastructures, archivists have taken, and are able to take, a range of approaches to collecting and caring for these records. Further, as this research revealed, the technologies and strategies subsequently used by archivists in order to collect records from within specific infrastructures also play a central role in shaping documentary heritage and collective memory. It is clear from this research that while

[516] Richard Pearce-Moses, "The Perfect and the Possible: Becoming a Digital Archivist," paper presented at the Conference of Inter-Mountain Archivists, May 12, 2006, www.lib.az.us/about/annualreports/2006/the.perfect.and.the.possible.becoming.a.digital.archivist.pdf.

socio-technical infrastructure clearly informs archival practice, there remains room for interpretation and judgement on the part of the archivists within these constraints.

7.2 Discussion

This dissertation began by asking how personal records have been traditionally defined and understood within the archival profession. Are they, as so many foundational theoretical texts have argued, fundamentally different from organizational records? And if they are, why is this so? Though Jackson's concept of broken world thinking would not be developed for many years after the publication of the Dutch Manual or the works of Jenkinson and Schellenberg, these archival theorists approached the task of defining personal archives in ways that can be effectively described and understood through the lens of Jackson's work. Through the act of defining personal records in the negative – by emphasizing how they break from archival theory – these archival scholars effectively shed more light onto the mechanisms of archival work more broadly, and in particular, the role of the archivist in shaping and bestowing value upon materials. Personal records have been defined by the ways in which they failed to comply with archival theory and practice; indeed, the idiosyncratic, unruly nature of personal records is one of their most defining characteristics. Because of this, personal archives have resisted tidy definitions. Archival scholars have advanced several frameworks for thinking about records beyond firm classifications of "personal" and "organizational." Two of these, the archival multiverse and activation, have proven incredibly in the context of this dissertation's research. The concept of the archival multiverse asks

archivists to consider how a single record can occupy myriad roles and functions.[517] Activation centers people in the work of conceptualizing records.[518] As this concept has been applied by Douglas and Mills, it asks archivists to consider to whom a record might be personal.[519] These concepts recognize that the meaning of records is not fixed, and that the role of the archivist is not passive. This dissertation concludes with an understanding of personal records as not only the records of private individuals, but as records that resist easy categorization and upend expectations of standardization. Personal records are not required for the purposes of accountability, as government records are understood to be. Instead, they provide valuable evidence of events as they were experienced by people; specifically, they provide insight into events as they were experienced by their creators. Taken individually, the provide "evidence of me;" collectively, they provide "evidence of us," as McKemmish has described this transformation.[520]

This perspective has provided valuable insight that shaped the investigation of this dissertation's other primary question, which is how the specific socio-technical infrastructures in which personal records are created and stored have affected archival approaches to collecting these materials. Each of the chapters that followed explored this question by examining three broad types of personal digital records: those created and stored on local devices; those created and stored on the open web; and those created and stored within networked, social platforms. As personal record-creation technologies have evolved, archivists have faced myriad challenges, both practical and theoretical, associated with acquiring and collecting records that have been created within emerging, evolving infrastructures. Additionally, archivists themselves have engaged new

[517] McKemmish and Piggott, "Toward the Archival Multiverse," 114.
[518] Ketelaar, "Tacit Narratives," 133.
[519] Douglas and Mills, "From the Sidelines to the Center," 260.
[520] McKemmish, "Evidence of Me," 186.

technologies, derived from a range of contexts beyond the archival profession, in order to collect, process, and provide access to these digital cultural heritage materials. This dissertation research has revealed that not only do the technologies with which records are created impact archival practice, but so too do the technologies used by archivists to work with those records.

Each of the collections examined in this dissertation supported arguments made in Chapter Three, which suggested that concepts such as "personal," "professional," and "organizational" are not themselves intrinsic qualities of records, but are often, instead, designations imposed by archivists. Evidence for this argument emerged in multiple distinct forms across collections. Some of the collections examined in this research, such as those of Rushdie, Sontag, Larsen, Lee, Mullaney, and Le Guin, contained the commingled records of all spheres of a creator's life, documenting their work, their relationships, and their daily thoughts and reflections. Other collection, including those in the #MeToo Collection, the Mormon Missionaries Collection, the Confederate Monument Collected Tweets, and the Zine Web Archive, have collected the records of many individuals in order to document a specific event, movement, or phenomenon. The GeoCities collections and the Twitter archive at the Library of Congress have been collected primarily on the basis of the platform in which they were created. These records are personal, but they are also shaped and informed by the socio-technical contexts of their creation, and their collection by archivists. Rather than attempting to position these collections as either personal or not personal, this research ultimately asks what is possible if we recognize their potential to possess multiple meanings to multiple stake holders?

Classification undoubtedly has its shortcomings. As this research has illustrated, what we call, or how we categorize, a body of records has a direct effect on how those records are treated – the rules, standards, and considerations that are applied to them. Thinking of records as an

archive of GeoCities or of Twitter, instead of thinking of them as the records of the individuals who created them, allows for the privileging of the terms of use established by platforms, rather than by people. Concepts like activation and the archival multiverse ask archivists to consider how a record might occupy multiple roles and meanings for multiple users at different points in time. These concepts are particularly useful for records created within commercial platforms, or even for those associated with specific hashtag movements. The Twitter archive, for example, is the body of public-facing records created within the commercial service Twitter; at the same time, it is made up of the records of thousands of private individuals. Though each of the collections examined in this dissertation contained the records of private individuals; however, in many instances, archival practice was focused on the structure of the records, rather than on the individuals who created them. This research affirmed that the infrastructure supporting personal records absolutely impacts the approaches taken by archivists.

This research has also conveyed the conditions of a profession that is working to continually adapt to a rapidly changing landscape, making the best possible use of the resources available. Indeed, these collections have gestured to the ways in which archivists draw upon and make use of pre-existing tools and technologies, some of which have been developed for archival purposes and some of which have come from other fields and disciplines. In Chapter Four, archivists incorporated tools from digital forensics, an area of specialty within the field of law enforcement. Used to process born-digital records, often acquired in obsolete file formats and stored on long-outdated computers and storage media, digital forensics tools and strategies have been instrumental in archival workflows. In Chapter Five, web crawling tools were used to capture web-based personal records in the majority of collections examined. In particular, the tools and services developed by the Internet Archive emerged essential infrastructure in this area. And in

Chapter Six, access to platform APIs enabled the collection of social media data at scale, making clear that the cooperation, whether tacit or implicit, of platform service providers continues to play a major role in social media archiving efforts. As with so many aspects of archival work, there is no single "right" path forward for archivists responsible for collecting personal digital archives of any kind. In the course of conducting this research, however, several themes emerged, which may prove informative or helpful to archivists faced with the myriad decisions that must be made when collecting or acquiring born-digital personal records.

7.3 Recommendations

This research was motivated by a desire to better understand the ways personal records have been understood and treated within the archival profession, as well as the effects on archival practice of the socio-technical infrastructures that support personal record creation today. Memory institutions are not, as Henniger and Scifleet have argued, "stable and unerring pillars of society," but are instead organizations, made up of people, who "actively test and revise their processes and procedures as the social, political, and technological landscape changes."[521] The collections explored throughout this dissertation support this notion. In collecting, preserving, and providing access to materials, archivists must make a series of decisions, based upon a variety of factors. These factors may include professional ethics, institutional priorities and policies, the preferences of record creators or donors, and the technological possibilities and constraints associated with born-digital archives. As personal computing technologies evolve, archival institutions are

[521] Henniger and Scifleet, "How are the New Documents of Social Networks Shaping our Cultural Memory?" 279.

periodically faced with the acquisition of materials for which they have not previously been responsible, including complete physical computers and computing environments. In some instances, these acquisitions are novel enough that it is difficult even to find comparable collections at other institutions. The acquisition of Salman Rushdie's personal computers at MARBL; the efforts to collect GeoCities at scale; and the Library of Congress' acquisition of the complete, public Twitter archive all stand as examples of acquisition for which archivists did not have a clear precedent for their work.

As discussed above, each of this dissertation's case studies depicted strategies employed by archivists faced with records situated within distinct, new infrastructures. Chapter Three demonstrated that personal records can be best characterized by their resistance to standardization and simple categorization. This is not a weakness of these records, nor a reason to exclude them from archives, but, instead, can be understood as a source of their value. Personal records are unique and specific to their creators, qualities that can make it difficult to establish hard and fast rules for working with them. However, this research has led to several recommendations that may help archivists as they acquire, and even build, collections that include personal records.

Consider the potential for records to be read as "personal"

This research has underscored the challenges – and flaws – of categorically defining records or collections as either "personal" or "not personal." Instead, archivists might ask themselves *how* and *to whom* records might potentially be understood or read as personal. Using conceptual models like activation and the archival multiverse, archivists might consider the multitude of interpretations and meanings of records, particularly when they are placed in new contexts, or made discoverable through new modes of access.

Include record creators in archival procedures

When personal records are locally stored, as with those discussed in Chapter Four, they are typically acquired as the result of an agreement between creator and collecting institution; only in select instances were similar agreements made with regard to personal records created and stored on the open web or in networked social platforms. It is imperative that archivists ask themselves why some record creators may determine the long-term fate of their digital records, and why others are not granted the same opportunity.

Ask what the potential privacy concerns and violations will be

Danielson has suggested that privacy violations are intrinsic to archival work. The question is not whether they will occur, but rather when. Contextual integrity can be employed as a useful framework for archivists who are attempting to establish how privacy can be violated or threatened by archival acquisition and access. Though contextual integrity alone cannot address all ethical concerns in archival practice, it establishes a baseline conceptualization of how the movement of records from personal custody to archival custody represents an impactful shift in context.

Engage in the critical study of personal computing technologies

As Chapters Four, Five, and Six explored, archival procedures for collecting personal digital records are shaped not only by the contents of the collection, or by the terms established by donors, but by the socio-technical environments in which those records are created. These chapters revealed that archival approaches to collecting personal digital records are likewise informed, to varying degrees, by both their creators and the environments in which they are created. In some instances, this research has revealed a tendency to collect personal records on the basis of what is technologically possible, rather than what is intended or desired by record creators. However, archivists must remember that there is no mandate to collect personal records, in spite of the informational and evidential value that they may offer.

Engage in the critical study of archival technologies

Additionally, it is clear that as a community, archivists (and archival educators) must think critically about the origins of the tools and methods we use, and how and where we apply them. As this research has revealed, archivists who work with personal records have been long lacked strong theoretical and practical frameworks specifically tailored to these materials. Archivists who work with personal digital records must undertake multiple acts of translation: translating methods designed for organizational records to personal records, and translating methods designed for paper records to suit digital records. In specific digital environments, archivists have drawn on tools from digital forensics, developed for law enforcement purposes; web archiving tools developed by technology companies; and APIs provided by commercial platforms. If archivists are to use tools developed in non-archival contexts, it is imperative that we ask how the values and ethics embedded within these tools align, or don't align, with archival values and ethics.

Develop specific appraisal policies for digital records

Throughout much of this research, there has been an underlying narrative about archival work as a means of rescuing what would otherwise be lost. Personal records, such as those discussed throughout this chapter, have the potential to provide meaningful evidence of lived experience. Advances in technology have not only made it possible for people to create more records; they have made it possible for archivists to collect more records. However, archivists have never saved everything; forgetting and remembering have always been part of the archival endeavor. Archivists who work with personal records must consider their own motivations for collecting these materials, but they must also consider what it means to collect them at scale. What is the relationship between archives and big data? Web and social media archiving tools, like the Wayback Machine and platform APIs, allow for the collection and retention of a massive volume

of records. This research suggests that there is still a need within the profession to consider the implications of collecting personal records as large datasets.

7.4 Future Research

This dissertation examined a variety of collections, containing personal records created in three distinct socio-technical environments. Within each of these chapters, unanticipated themes emerged alongside the dissertation's intended thematic structure. Each of these themes has inspired new questions, and it is my intention to address these more thoroughly in future research projects. Where the structure of this dissertation has been broad, it has helped me to identify specific issues and phenomena that warrant further, deeper focus in the future. Indeed, each of the chapters in this dissertation has generated interest in new questions and areas of focus that specifically address personal records within these socio-technical infrastructures.

In Chapter Four, the adoption of digital forensics tools and methods played a significant role in the processing of born-digital materials acquired on physical storage media. Specifically, I am interested in the use of digital forensics tools for the purpose of recovering deleted files in personal digital archives, and the extent to which such recovery is common in archival practice. This research has suggested that a comparison of the values, particularly with regard to discovery and privacy, upheld by digital forensic science and archival science could be of use to archivists who use these tools to process personal digital records acquired on physical storage media.

In Chapter Five, the pervasive role of the Internet Archive in web archiving came into focus. While I had previously been well aware of the Internet Archive's Wayback Machine and its own collections and collecting practices, I had failed to see the full extent of its influence

throughout the professional web archiving landscape. Many archivists have been quick to point out that the Internet Archive is not an archival repository as they are traditionally defined within the field, but is instead a non-profit digital library that shares some common interests and objectives with the archival community. However, there has not yet been an independent investigation into the degree to which archival institutions rely upon the tools and services the Internet Archive provides. This seems an incredibly fruitful area for further investigation.

And in Chapter Six, it became clear how directly the Terms of Service and Privacy Policies of social media platforms influences collecting practices. In order to lawfully collect the records created within these platforms, archivists are compelled to conform to the terms of use that they have established. Often, they often must also comply with the parameters set by platform APIs. Moving forward, I am interested in thinking more deeply about how the expectations of service providers intersect with those of individual record creators. How do archivists negotiate these potentially conflicting interests? How do these socio-technical environments further complicate archival notions of the personal and the organizational?

Throughout each of these areas, I remain interested in the intersection of archival ethics and technology, and the evolving relationships between record creator, archivist, and service provider. The collections studied throughout this dissertation have revealed a wide range of approaches taken to collecting personal digital records. In some instances, the fear of potential loss of digital cultural heritage appears to be a driving motivation for collection; in others, relationships between record creators and archivists establish the terms under which collections are acquired, preserved, and made accessible. Although the Twitter archive at the Library of Congress remains a lone example of a social media platform donating its entire public archive to a memory institution – and although this collection has been deemed a failure in many respects – the possibility of future

agreements between service providers and archives does not seem to be out of the question. As long as personal records are created in these socio-technical systems, commercial service providers will play a significant role in the long-term preservation of our digital cultural heritage.

Appendix A Collection Summaries

Appendix A consists of summaries of each of the collections discussed in this dissertation. These summaries describe the scope and contents of each collection, as well as the institutional context in which the collections have been preserved and made accessible. They are arranged according to the chapter in which they were discussed, with each chapter's collections organized alphabetically.

Appendix A.1 Chapter Four Collection Summaries

Appendix A.1.1 Rafael Fajardo Digital Materials

The digital records of artist, designer, and scholar Rafael Fajardo reside at the Media Archaeology Lab (MAL) at the University of Colorado Boulder.[522] The collection's documentation identifies Fajardo as the donor of these materials to the MAL. At the time of this writing, the materials remain unprocessed. Selected items are listed within the public MAL catalog, which is organized according to physical medium, rather than donor. Fajardo's materials are stored within a single box on a shelf within the lab. The collection contains an impressive variety of storage

[522] Media Archaeology Lab, https://mediaarchaeologylab.com/. In May 2019 I completed a research residency at the MAL. My objective for the residency was to gain more hand-on experiences with the personal computers and other tools used for early personal digital record creation. It was entirely due to the expertise and generosity of the MAL staff that Fajardo's materials were brought to my attention, and I am indebted to them for making the connection. Fajardo currently teaches at the University of Denver, and his portfolio can be viewed at https://www.rafaelfajardo.com/.

media, SCSI Mirror SyQuest 44 MB removable disk cartridges, Iomega 1 GB Jaz Drives, Apple CD 150 and 600e, and 3.5" floppy disks. These storage devices contain design files and documentation from Fajardo's artistic endeavors; HTML files from websites created by Fajardo and his collaborators; and assignments completed by Fajardo's students, as well as his comments on their work. The records stored on the SyQuest disk cartridges could not be accessed during my time at the MAL due to lack of the equipment required to read the disks. Minimal information about their contents was derived from handwritten notes on labels affixed to the disks.

Fajardo's collection stands as an example of how some digital materials continue to perpetuate the "shoebox" metaphor sometimes used to describe personal archives. The materials remain stored within the box, where they can be accessed by researchers; none of the digital records included within the collection have been migrated to more contemporary storage media, nor have they been subject to any other preservation activities commonly associated with born-digital records. Further, this collection is noteworthy as an example of one's personal digital archive entering into the custody of an institution that is not primarily an archives, but instead a research lab designed to facilitate and support "cross-disciplinary experimental research and teaching using still functioning media from the past."[523] As such, it reflects the role that non-archival institutions play in the long-term stewardship of digital cultural heritage.

[523] "What," Media Archaeology Lab, https://mediaarchaeologylab.com/about/what/.

Appendix A.1.2 Deena Larsen Collection

In 2007, the Maryland Institute for Technology in the Humanities (MITH) at the University of Maryland acquired the archival materials of new media and hypertext writer Deena Larsen.[524] The Deena Larsen collection provides documentation of Larsen's work and life, the work of Larsen's contemporaries, and of computing history more broadly. The collection thus constitutes not only a record of Larsen's own activities and output, but of the creative community and social network in which she has worked. The contents of the collection date primarily from the early 1980s through the early 2000s, and include both digital and non-digital formats. Larsen's collection contains "manuscripts, newspaper clippings, books, comics, manuals, notebooks, syllabi, catalogs, brochures, posters, conference proceedings, ephemera, and a shower curtain."[525] The shower curtain is an analog representation of the linked records found within her most well-known work of electronic literature, *Marble Springs*.[526]

Larsen's collection is unique in that its contents being primarily digital, with non-digital materials in the minority.[527] Included among these digital materials, are over 800 diskettes and nearly 100 CD-ROMs and Zip disks.[528] These storage devices contain files in an assortment of types and formats. In addition to these storage devices are the personal computers used by Larsen early in her career, including five Mac Classics, two Mac SEs, and one Mac Plus.[529] Larsen's

[524] Matthew Kirschenbaum, "About the Deena Larsen Collection," 2009, http://mith.umd.edu/larsen/about/about.
[525] Kirschenbaum, "About the Deena Larsen Collection," https://mith.umd.edu/larsen/about/about/.
[526] Deena Larsen, "Marble Springs shower curtain," The Deena Larsen Collection at the Maryland Institute for Technology in the Humanities, (n.d.). Accessed http://mith.umd.edu/larsen/items/show/42.
[527] In the collections examined in Chapter Four, digital materials more frequently constitute only a small portion of the overall contents.
[528] An inventory of the collection contents can be downloaded as an Excel spreadsheet from the collection website. "The Collection Finding Aid," The Deena Larsen Collection, https://archive.mith.umd.edu/larsen/about/about/index.html.
[529] Details and images of Larsen's Mac Classic II can be seen on the collection website. "Mac Classic II," https://archive.mith.umd.edu/larsen/items/show/159/index.html.

collection is noteworthy not only because it allows users to access the original hardware and software once used Larsen herself, but because it serves as an example of the resources that are required in order to preserve complex digital objects, such as works of interactive fiction or complete personal computers. Together, these resources are illustrative of the interdependencies among hardware, software, and record.

Appendix A.1.3 Toni Morrison Papers

In 2014, representative of Princeton University announced that the papers of Nobel Laureate, author, and former Princeton faculty member Toni Morrison had been acquired by the University.[530] By 2016, a major portion of the collection had been processed and made accessible to researchers.[531] Morrison's collection contains "about 180 linear feet of research materials documenting the author's life, work, and writing methods," according to Princeton University Library curator of manuscripts Don Skemer.[532] The collection is primarily comprised of paper and other non-digital records, including manuscripts, drafts of novels, teaching materials, correspondence, and photographs, but included among these materials are over 150 floppy disks.[533] These disks were home to both personal and professional born-digital records, including drafts of academic lectures, correspondence, personal administrative documents, and early drafts of

[530] "Toni Morrison Papers to Reside at Princeton," Princeton University News, October 17, 2014, https://www.princeton.edu/news/2014/10/17/toni-morrison-papers-reside-princeton.
[531] "Toni Morrison Papers Open for Research," Princeton University Library, June 8, 2016, https://library.princeton.edu/news/general/2016-06-08/toni-morrison-papers-open-research.
[532] "Toni Morrison Papers to Reside at Princeton."
[533] Colon-Marrero and Hughes, "Toni Morrison's Born-Digital Material," *Mudd Manuscript Library Blog*, August 26, 2015. https://blogs.princeton.edu/mudd/2015/08/toni-morrisons-born-digital-material/.

Beloved that had previously been believed to be lost.[534] These files are primarily text-based, created with several word processing software programs, including WordPerfect and Microsoft Word. Additional digital materials can be found in the collection's Audiovisual Materials series, which contains digital photographs, both digitized and born digital. The documents retrieved from Morrison's floppy disks have proven to be of particular interest as a case study in using digital forensics tools and methods in the processing of born-digital records. Documentation created by processing archivists provides valuable insight into the technical challenges of transferring these records from obsolete media into archival management systems for long term preservation and access.[535]

Appendix A.1.4 Salman Rushdie Papers

In 2006 novelist Salman Rushdie joined Emory University's faculty as Distinguished Writer and Residence and deposited his papers with the Manuscript, Archives, and Rare Book Library (MARBL), located with the university's Robert W. Woodruff Library.[536] The hybrid collection consists of both papers and born-digital materials from Rushdie's career and personal life, including "a large quantity of correspondence with a wide literary circle, materials documenting Rushdie's life under the fatwa, notebooks and journals maintained since 1973, photographs, and other related personal and literary papers."[537]

534 Toni Morrison Papers, "Floppy Disks," https://findingaids.princeton.edu/collections/C1491/c4181; "Toni Morrison Papers Open for Research," Princeton University Library.
535 Elena Colon-Marrero and Allison Hughes, "Toni Morrison Collection," BitCurator Consortium Case Studies, https://www.bitcuratorconsortium.org/case-studies/toni-morrison-collection. See also Jarret Drake, "The University Archives and its Focus on Fixity," Mudd Manuscript Library Blog, October 10, 2014, https://blogs.princeton.edu/mudd/2014/10/the-university-archives-and-its-focus-on-fixity/.
536 "Emory Acquires Rushdie Archive," *Georgia Library Quarterly* 43, no. 4 (2007): 22.
537 Emory Acquires Rushdie Archive," 22.

MARBL archivists noted that "while the collection consists of over one hundred linear feet of traditional archival material, such as journals, correspondence, and manuscript writings, the reason that this collection stands out from the rest of those housed at MARBL is its large born-digital component."[538] Indeed, as Erika Farr and Dorothy Waugh have observed, at the time of acquisition, the collection "included a nearly unprecedented large born-digital component, including the computers upon which Rushdie wrote most of his novels."[539] In addition to the physical computers and storage media themselves, the collection contained extensive digital records, including "notes and drafts of Rushdie's writings, daily calendars, correspondence, personal and financial files, games, photographs, and downloaded web pages."[540] The majority of Rushdie's born-digital materials date from between 1992 and 2006. [541] All digital materials were received by archivists on local physical storage media.

In order to address "the new challenges and issues involved in preserving the born-digital material, as well as making it accessible to researchers in an innovative and responsible way that incorporated both donor concerns and user expectations," archivists at MARBL formed a Born-Digital Archives (BoDAR) working group.[542] The group ultimately decided to take a multi-pronged approach to working with the collection, which included not only copying and migrating files into their archival management system, but emulating Rushdie's the desktop environment of one of Rushdie's personal computers, allowing researchers to access the author's records through the same interface that he used to create and use them. [543] These materials are available only within

[538] Carroll et al. "A Comprehensive Approach to Born Digital Archives," *Archivaria,* 63.
[539] Erika Farr and Dorothy Waugh, "Salman Rushdie Archive," BitCurator Consortium Case Studies, https://www.bitcuratorconsortium.org/case-study/salman-rushdie-archive.
[540] Carroll, et al., "A Comprehensive Approach to Born-Digital Archives," 64.
[541] Carroll, et al., "A Comprehensive Approach to Born-Digital Archives," 64.
[542] Carroll et al. "A Comprehensive Approach to Born Digital Archives, 65.
[543] Patricia Cohen, "Fending Off Digital Decay, Bit by Bit; Dan Rockmore, "Searching Through the Salman Rushdie Archives."

MARBL's reading room, where a dedicated computer can be used to access the emulation. The inclusion of Rushdie's personal computers, and the MARBL archivists' decision to provide access to the author's desktop environment through emulation, received dedicated attention in the mainstream media.[544]

Appendix A.1.5 Susan Sontag Papers

The archives of writer, critic, and activist Susan Sontag were acquired by the Department of Special Collections at the University of California, Los Angeles' Charles E. Young Research Library in 2002. The collection was reportedly purchased for $1.1 million, an acquisition that was made possible by funds from an anonymous UCLA alum.[545] The Sontag collection consists of 132 linear feet of material, housed in 264 manuscript boxes, 67 oversize boxes, and one map folder.[546] These aggregated records contain Sontag's personal and professional records, including correspondence, teaching materials, notes and drafts of both published and unpublished works, drawings and stories she created as a child, and personal and academic records created during her years as a student.[547]

Sontag's born-digital materials were acquired as a part of the 2012 addition to the writer's papers. Born-digital materials were processed first in 2014, and again in 2018 and 2019 by staff in UCLA's Center for Primary Research and Training.[548] The digital portion of the collection dates primarily from between 1995 and 2002. The 7.7 gigabytes of digital materials, composed of 64,461

544 Rockmore, "Searching Through the Salman Rushdie Archives;" Cohen, "Fending Off Digital Decay, Bit by Bit."
545 Avins, "UCLA Buys Sontag's Archive."
546 Biber and Luker, "Evidence and the Archive: Ethics, Aesthetics, and Emotion."
547 Biber and Luker, "Evidence and the Archive: Ethics, Aesthetics, and Emotion."
548 Susan Sontag Papers, http://www.oac.cdlib.org/findaid/ark:/13030/kt2489n7qw/admin/#did-1.2.1

digital files, are arranged and described as a single series within the collection. An item-level inventory of digital files, totaling 38 pages, has been included in the collection finding aid.[549] The Scope and Content note for this series offers a more succinct overview of its contents, which include "agendas, calendars and itineraries; notes; financial, medical and travel documents; lists of favorite words, books, and films; lists of hotels, restaurants, museums and other destinations, arranged geographically; music files; and reproductions of various genres of artwork."[550] The born-digital materials within this collection bear some similarities to those of Salman Rushdie at Emory University, including the fact that both collections contain complete computing environments, and both have received substantial media attention.[551]

Appendix A.2 Chapter Five Collection Summaries

Appendix A.2.1 The Archive Team GeoCities Collection

In 2009, Yahoo announced that it would cease to support the web hosting service GeoCities, permanently deleting all extant websites created with the once-popular web hosting service.[552] At the height of its popularity GeoCities, which was founded in 1994 as Beverly Hills

[549] Susan Sontag papers.
[550] Susan Sontag papers.
[551] See Allison P. Davis, "Susan Sontag, Sephora Shopper," *The* Cut, October 27, 2014, https://www.thecut.com/2014/10/susan-sontag-sephora-shopper.html; Benjamin Moser, "In the Sontag Archives," *The New Yorker*, January 30, 2014, https://www.newyorker.com/books/page-turner/in-the-sontag-archives; Jeremy Schmidt and Jacquelyn Ardam, "On Excess: Susan Sontag's Born-Digital Archive," *Los Angeles Review of Books*, October 26, 2014, https://lareviewofbooks.org/article/excess-susan-sontags-born-digital-archive/; "The New Age: Leaving Behind Everything, or Nothing At All," *All Things Considered*, April 9, 2014, https://www.npr.org/sections/alltechconsidered/2014/04/09/300614977/the-new-age-leaving-behind-everything-or-nothing-at-all.
[552] "Yahoo Sets the Date of GeoCities' Death," *PC Mag*, July 10, 2009, https://www.pcmag.com/archive/yahoo-sets-the-date-of-geocities-death-242171.

Internet, had been the third most frequently-visited site on the World Wide Web. In 1999, it was purchased by Yahoo, and over the decade that followed, it fell out of popular use.[553] When it was announced that GeoCities would be deleted in October of 2009, a group of self-described "rogue archivists" known as the Archive Team launched a grassroots effort to capture and preserve as many GeoCities websites as they could in the time that they had. As Archive Team states on their website, "to not at least have the option of browsing these old sites would be a loss of the very history of the web from the side of the people who came to know it, not the designers who descended upon it. For that reason, Archive Team thinks GeoCities is worth saving."[554]

The Archive Team's GeoCities collection stands out from other efforts to capture and preserve GeoCities in that it has not hosted the preserved content as a collection of browsable websites, but instead as a 641.4 GiB torrent download on the download site The Pirate Bay.[555] Users of this archive download the torrent and then store their own, downloaded copy on their own storage systems for indefinite future use. The result of this system is an unknown number of copies of the GeoCities collection within the personal computing devices and accounts of those who have downloaded it throughout the past decade.

Appendix A.2.2 GeoCities Special Collection (Internet Archive)

The Internet Archive's GeoCities Special Collection 2009 is home to another collection of websites created with the GeoCities service.[556] Like the Archive Team GeoCities collection, this

[553] Leena Rao, "Yahoo Quietly Pulls the Plug on GeoCities," *TechCrunch*, April 23, 2009, https://techcrunch.com/2009/04/23/yahoo-quietly-pulls-the-plug-on-geocities/.
[554] Archive Team, "Glorious History," GeoCities, https://www.archiveteam.org/index.php?title=GeoCities.
[555] "GeoCities – The PATCHED Torrent," The Pirate Bay, uploaded April 29, 2011, https://thepiratebay.org/description.php?id=6353395.
[556] "GeoCities Special Collection 2009," Internet Archive, https://archive.org/web/geocities.php.

one was built in 2009, in the time between Yahoo's announcement and the deletion of GeoCities. The Internet Archive's Wayback Machine was used to capture a sample of GeoCities websites through "several special deep collection crawls," and successfully captured "specific sites nominated by the public, over the last few months that GeoCities was in operation, to help make our archive of GeoCities sites as deep and thorough as possible."[557] The collection website acknowledges the work of Archive Team volunteers, who contributed URL surveys to the Internet Archive's project in addition to their own web archiving efforts.

User access to the sites in the GeoCities collection is functionally similar to that of other sites captured with the Wayback Machine. In order to visit an archived version of a site, the user must enter the complete URL and navigate directly to the site; browsing and keyword searching are not available. However, using the URL lists made available elsewhere online, a user may be able to successfully locate the URL for a specific URL they are seeking, or isolate the addresses of websites created within specific neighborhoods.[558] Two of the three URL lists cited by Archive Team, including the one used by the Internet Archive, are no longer available online, even in archived from.[559] However, the "GeoCities biglist" remains on Archive Team's website, and can be used to source URLs that can be retrieved from the Internet Archive's collection.[560]

[557] Internet Archive, "Saving a Historical Record of GeoCities," https://archive.org/web/geocities.php.
[558] In addition to being among the most popular early website-building services, GeoCities has been remembered for its unique "neighborhood" structure. When creating websites, users chose sub-domains based on topics addressed through their websites. Ian Milligan has written about the role that the neighborhood structure played in fostering a sense of community on the World Wide Web in the days before social networking and other Web 2.0 functionalities. See "Welcome to the Web: The Online Communities of GeoCities during the Early Years of the World Wide Web," in *The Web as History: Using Web Archives to Understand the Past and the Present,* ed. Niels Brügger and Ralph Schroeder (London: UCL Press, 2017): 137-158.
[559] "GeoCities URL Lists," Archive Team, https://www.archiveteam.org/index.php?title=GeoCities_URL_Lists.
[560] "GeoCities biglist," Archive Team, https://www.archiveteam.org/index.php?title=Geocities_biglist.

Appendix A.2.3 The Zine Web Archive

Web archiving at the Library of Congress began with the MINERVA (Mapping the Internet Electronic Resources Virtual Archive) project in 2000, and continues today under the title of the Library of Congress Web Archives.[561] Websites are selected for inclusion in the archives by subject experts known as Recommending Officers, based on the Library's collecting policies and subjects relevant to its mission.[562] The Zine Web Archive was established by the Library of Congress's Digital Collections department in 2018, "to supplement the physical zine collection" held by the Library.[563] A zine is an informal, underground publication with roots in social and political activism and punk music.[564] While zines originated as a paper-based medium, they have more recently been created in digital form. Online communities have also arisen to support the distribution and discussion of both print and digital zines. The Library of Congress's Zine Web Collection documents these web-based aspects of zine culture. The collection contents include archived versions of web zines, review websites, and community forums. The scope and content of the web archive reflects the collecting priorities of the print zine collection, which include "zines made by people of color, women, immigrants, and LGBTQ+ and transgender and gender non-conforming individuals and organizations. Subjects and perspectives which have been traditionally underrepresented in mainstream media (and therefore libraries) are also a high priority for

561 Jeffrey Garret, "An Evaluation of Web Archiving Programs in the U.S. Relevant to International and Area Studies," Center for Research Libraries Report, February 2019, https://www.crl.edu/sites/default/files/reports/Garrett%20Report%202019.pdf.
562 "Web Archiving Frequently Asked Questions," Library of Congress, https://www.loc.gov/programs/web-archiving/about-this-program/frequently-asked-questions/.
563 "Zine Web Collection," Library of Congress, https://www.loc.gov/collections/zine-web-archive/about-this-collection/.
564 "What is a Zine?" University of Texas Libraries, https://guides.lib.utexas.edu/c.php?g=576544&p=3977232.

collection."[565] As self-published, non-commercial materials, zines allow for "an unprecedented freedom of expression, and as such, these materials are incredibly valuable primary source materials."[566]

The collection website states that while websites in the archive are "selected by Recommending Officers and Subject Specialists from across the Library… this collection is primarily curated by the Collection Specialist for Women's Gender, and LGBTQ+ Studies."[567] At the time of this writing, the collection includes 15 unique websites.[568] These are captured monthly or yearly, with a smaller selection of sites captured "quarterly, twice-yearly, or once."[569] This collection, while small, is still being actively built and expanded.

Appendix A.2.4 Katie Lee Collection

The Katie Lee Collection is held by the Cline Library at Northern Arizona University.[570] The collection is hybrid, containing both digital and non-digital materials that document the personal and professional activities of Lee, a singer, songwriter, author, and environmentalist from Arizona.[571] The collection spans nearly 30 linear feet of textual materials in addition to audiovisual materials and an extensive collection of photographs, slides, and negatives. Digital materials make

[565] "About this Collection – Zine Web Collection," Library of Congress.
[566] "About this Collection – Zine Web Collection," Library of Congress.
[567] "About this Collection – Zine Web Collection," Library of Congress.
[568] "Collection Items – Zine Web Collection," Library of Congress, https://www.loc.gov/collections/zine-web-archive/.
[569] "About this Collection – Zine Web Collection," Library of Congress.
[570] Katie Lee Collection, 1719-2019, http://www.azarchivesonline.org/xtf/view?docId=ead/nau/lee_katie.xml&doc.view=print;chunk.id=0.
[571] "Abstract," Katie Lee Collection, 1719-2019.

up a small portion of the collection. Included among these materials are three websites, arranged in their own record group within the collection.

The websites in Record Group 6 date from between 2008 and 2019. They include Katie Lee's personal website, which was created and maintained by Lee until its accession by the Library's Special Collections and Archives in 2017, as well as two online exhibits about Lee that have been created by the library's staff. Lee's personal website, www.katydoodit.com, was built with the open source software WordPress. It contains embedded audio and video, writing by Lee, transcriptions of interviews conducted with her throughout her career, a collection of her newsletters, and general biographical information. Its homepage has been updated to include the announcement of Lee's death in 2017.[572] The exhibit "Naked Truth," was created within Lee's personal website. The other exhibit, "Full Circle: The Life & Legacies of Katie Lee," was created in a separate WordPress website, and is also hosted by the library.[573]

Appendix A.2.5 Mormon Missionaries Collection

The Mormon Missionary Collection at Brigham Young University, located in the university library's L. Tom Perry Special Collections, documents Mormon missionary experiences from the 1830s to the present day.[574] The collection is "one of the premier collections of historical documents related to the history of missionary work in The Church of Jesus Christ of Latter-day Saints, including journals, letters, photographs, scrapbooks, autobiographies, and oral histories."[575]

572 https://www.katydoodit.com/
573 Katie Lee Collection, https://library.nau.edu/speccoll/exhibits/katielee/
574 "Mormon Missionary Collections," Brigham Young University Library, https://lib.byu.edu/collections/mormon-missionary-collections/.
575 "Mormon Missionary Collections," Brigham Young University Library.

A selection of the early, handwritten diaries have been digitized and made accessible to a larger, off-site audience. According to the collection website, selections for digitization were made from the 575 volumes, documenting the lives of over 220 diarists.[576] As documentation of missionary life expanded to include blogging, the Mormon Missionary Collection incorporated web archiving into its collecting strategy in order to capture and preserve these born-digital records.

The Web Archive component of Mormon Missionary Collection includes 5,331 blogs created by missionaries during their periods of service.[577] Its contents date from between 2011 and 2014, and have been collected by BYU since March 2013. The blogs and websites were captured with Archive-It, a subscription-based web archiving service provided by the Internet Archive. Archive-It provides a suite of tools and hosting packages for archives and other collecting institutions through a user-friendly, web-based application.[578] The resulting collection can be searched by keyword or browsed by creator, contributor, subject, date, or language.

Appendix A.3 Chapter Six Collections

Appendix A.3.1 #MeToo Collection

The "Me Too" movement, often written as the #MeToo movement, was founded in 2006 by activist and educator Tarana Burke, to "help survivors of sexual violence, particularly Black women and Girls, and other young women of color from low wealth communities, find pathways

[576] "About – Mormon Missionary Diaries," Brigham Young University, https://lib.byu.edu/collections/mormon-missionary-diaries/about/.
[577] "Mormon Missionary Collection," Archive-It, https://archive-it.org/collections/3609.
[578] Archive-It, https://archive-it.org/blog/learn-more/

to healing."[579] In 2017, following allegations against Harvey Weinstein, the hashtag #MeToo spread virally on social media platforms, and on Twitter in particular.[580] In 2018, Harvard University's Schlesinger Library announced the establishment of its #MeToo Collection, which would collect records related to the movement and associated with the hashtag. According a statement made by Amanda Strauss, Manager of Special Projects and Digital Services at Schlesinger Library, in 2018, "the digital footprint of #MeToo in the past year measures more than 19 million English-language Twitter posts and thousands of news articles and personal testimonials."[581] In order to ensure that these records, many of which were created in proprietary social media platforms, remained accessible to "scholars ranging from historians to data scientists," Strauss continued, "the Schlesinger Library, with support from a generous S.T. Lee Innovation Grant from Harvard Library, has started a large-scale project to comprehensively document #MeToo."[582]

In order to service its stated goal of documenting the "digital footprint of the #metoo movement and the accompanying political, legal, and social battles in the United States," the collection includes "social media, news articles, statements of denial and/or apology, Web-forum conversations, legislation, lawsuits, statistical studies, Fortune 500 companies' employment manuals, hashtags related to #metoo, and more."[583] Materials are identified and collected through a variety of methods. Archivists at Schlesinger track a series of hashtags, collecting each social

[579] "History and Vision," Me Too, https://metoomvmt.org/about/#history.
[580] Nadia Khomami, "#MeToo: How a Hashtag Became a Rallying Cry Against Sexual Harassment," *The Guardian*, October 13, 2017, https://www.theguardian.com/world/2017/oct/20/women-worldwide-use-hashtag-metoo-against-sexual-harassment.
[581] Amanda Strauss, "#MeToo: A Glimpse into the Digital Vault," Schlesinger Newsletter, Fall 2018, https://www.radcliffe.harvard.edu/news/schlesinger-newsletter/metoo-glimpse-digital-vault.
[582] Strauss, "#MeToo."
[583] "Collection Goal & Scope Statement," #MeToo Digital Collection, https://www.schlesinger-metooproject-radcliffe.org/collection.

media post that uses each tag.[584] Users may also submit or recommend materials to be included.[585] At the time of this writing, the Twitter data in the collection is not publicly accessible.[586] Users may, however, access the 1,108 archived websites included in the collection, which have been captured and made accessible with Archive-It's subscription web archiving service.[587]

Appendix A.3.2 Thomas S. Mullaney Papers

The Stanford University Archives collects, as part of its mission, the papers of its faculty.[588] Among these are the papers of Thomas S. Mullaney, a Professor of Chinese History at Stanford University and Guggenheim Fellow.[589] The collection contains 23.5 linear feet material, including audio recordings, journals, manuscripts and drafts, and academic documents from Mullaney's youth and career, which remain closed to researchers until 2060.[590] Included in the collection are Online Materials, which have been made available through Stanford's Archive-It repository.

Among the six items in this portion of the collection is Mullaney's Twitter profile, which has been captured twice in December 2018.[591] The Archive-It records show the contents of Mullaney's Twitter profile as they appeared at the time of capture. Pages, profiles, and tweets that

[584] "Collection," #MeToo Digital Collection, https://www.schlesinger-metooproject-radcliffe.org/collection.
[585] "Contribute to the Collection," #MeToo Digital Collection, https://www.schlesinger-metooproject-radcliffe.org/contribute.
[586] The collection page states that "Twitter data will be available for research in 2020." "Access the Collection," #MeToo Digital Collection, https://www.schlesinger-metooproject-radcliffe.org/access-the-collection.
[587] #metoo Web Archives Collection, Archive-It, https://archive-it.org/collections/10866.
[588] "What We Collect," Stanford University Special Collections and University Archives, https://library.stanford.edu/spc/university-archives/about-archives/what-we-collect.
[589] "Tom Mullaney," Stanford University History Department, https://history.stanford.edu/people/tom-mullaney.
[590] Thomas S. Mullaney Papers SC1435SC1435, Stanford University Archives, http://pdf.oac.cdlib.org/pdf/stanford/uarc/sc1435.pdf.
[591] "Https://twitter.com/tsmullaney," Thomas S. Mullaney Web Archive (Stanford University Archives), https://wayback.archive-it.org/10862/*/https://twitter.com/tsmullaney.

are linked from Mullaney's profile have not been crawled, and are thus unavailable to viewers of the archived version of his profile. It is not possible, for example, to navigate to the tweets Mullaney has liked, his followers, or the accounts he follows. Instead, users are able to scroll through the main profile feed, viewing Mullaney's original tweets, beginning with his pinned tweet, announcing an award received for his book *The Chinese Typewriter: A History*, and his most recent tweet at that time, made on November 30, 2018. The archived version of the profile provides access to one facet of Mullaney's social media presence, but limits access to the full interactive functionality of the service.

Appendix A.3.3 Twitter Archive

On April 15, 2010 the Library of Congress made announced that it had acquired the social media platform Twitter's complete public archive, which dated back to 2006.[592] Moving forward, the announcement continued, the Library would collect all public tweets on an ongoing basis. James H. Billington, Librarian of Congress at that time, noted that the collection had the potential to fuel future research "into our contemporary way of life," as well as the potential to provide "detailed evidence about how technology based social networks form and evolve over time."[593] Some questioned the value of the acquisition, but as the Library's Director of Communications Gayle Osterberg later argued, the Twitter archive was well-suited to the Library's mission and collecting goals:

[592] Matt Raymond, "Twitter Donates Entire Tweet Archive to Library of Congress," News from the Library of Congress, April 15, 2010, https://www.loc.gov/item/prn-10-081/twitter-archive-to-library-of-congress/2010-04-15/.
[593] Raymond, "Twitter Donates Entire Tweet Archive to Library of Congress."

> "An element of our mission at the Library of Congress is to collect the story of America and to acquire collections that will have research value. So, when the Library had the opportunity to acquire an archive from the popular social media service Twitter, we decided this was a collection that should be here."[594]

The collection grew at a rapid pace; by 2013, the Library's had acquired approximately 170 billion tweets.[595] Periodic transfers of data from Twitter to the Library continued until 2017, when Library announced that it would no longer collect all public tweets, as it had been since the 2010 agreement. Instead, they would cease comprehensive collecting on December 31, 2017, opting instead for a selective appraisal strategy, consistent with their general collecting policies for web-based materials.[596] Moving forward, they wrote, "the tweets collected and archived will be thematic and event-based, included events such as elections, or themes of ongoing national interest, e.g. public policy."

In the decade since the initial agreement between the Library and Twitter, the collection has yet to be made accessible to researchers. In the 2017 update on the project, the Library indicates that the collection "will remain embargoed until access issues can be resolved in a cost-effective and sustainable manner."[597]

[594] Gayle Osterberg, "Update on the Twitter Archive at the Library of Congress," Library of Congress Blog, January 4, 2013, https://blogs.loc.gov/loc/2013/01/update-on-the-twitter-archive-at-the-library-of-congress/.
[595] "Update on the Twitter Archive at the Library of Congress," White Paper, January 2013, https://www.loc.gov/static/managed-content/uploads/sites/6/2017/02/twitter_report_2013jan.pdf.
[596] The Library of Congress' Collections Policy Statements are available at https://www.loc.gov/acq/devpol/.
[597] Gayle Osterberg, "Update on the Twitter Archive at the Library of Congress," Library of Congress Blog, December 26, 2017, https://blogs.loc.gov/loc/2017/12/update-on-the-twitter-archive-at-the-library-of-congress-2/.

Appendix A.3.4 Confederate Monument Protests Collected Tweets

This collection, held by the University of North Carolina at Chapel Hill's University Archives, contains tweets primarily associated with the Twitter hashtags "#SilenceSam and #SilentSam, as well as a smaller sampling of tweets associated with related hashtags. These tweets were created in support of the movement to remove the Confederate monument, known as Silent Sam, on the UNC Chapel Hill campus.[598] The Collection Overview notes that the collection "represents only a snapshot of tweets related to significant protests and events from August 2017 through December 2017, May 2018, August 2018 through September 2018, December 2018 through May 2019, August 2019, December 2019, and February 2020."[599]

The collection itself does not contain the full text of these tweets, as that degree of access would violate Twitter's Terms of Service. Instead, "the collection is made available as files containing tweet identifiers (tweet ids) for all collected tweets. The identifiers can be used to query the Twitter API and gain access to full tweet data."[600] Twarc may be used to "hydrate" the tweet identifiers, another way of gaining access to their full contents. Twarc is a tool for archiving Twitter JSON data, developed by the Documenting the Now project team.[601] Using the provided tweet identifiers, Twarc can work with Twitter's API to "*hydrate* the data, or to retrieve the full JSON for each identifier."[602] This allows researchers who wish to use the tweets in this collection to verify their contents, as well as to further contextualize them with metadata provided by Twitter.

598 "Collection Overview," University of North Carolina at Chapel Hill Confederate Monument Protests Collected Tweets, https://finding-aids.lib.unc.edu/40486/.
599 "Collection Overview," University of North Carolina at Chapel Hill Confederate Monument Protests Collected Tweets.
600 "Restrictions to Access," University of North Carolina at Chapel Hill Confederate Monument Protests Collected Tweets.
601 "Twarc," Documenting the Now GitHub, https://github.com/DocNow/twarc.
602 "Twarc," Documenting the Now GitHub, https://github.com/DocNow/twarc.

Appendix A.3.5 Ursula K. Le Guin Papers

Author Ursula K. Le Guin had a longstanding relationship with the University of Oregon, and began depositing her papers with the University Libraries in 1980, according to the university manuscripts librarian Linda Long, who worked directly with Le Guin and her collection for over 20 years.[603] Le Guin's papers at the University of Oregon date from approximately 1930 through 2018, the year of Le Guin's passing.[604] The 140.25 linear feet of materials in the collection include "correspondence, literary works, legal and financial files, public appearances and publicity materials, personal papers, photographs and artwork, audiovisual material, website and social media, and writing of others," according to the collection finding aid.[605]

Series VIII: Website and Social Media within the collection consists of Le Guin's personal website and social media accounts, which include a blog, Facebook account, and Instagram account. These materials have been collected and made accessible through the subscription archiving service Archive-It.[606] Le Guin's public-facing Facebook and Instagram profiles have each been captured one time, on July 5, 2018 and July 2, 2018 respectively. By the time of each capture, both of these social media accounts had become memorial accounts, managed by Le Guin's estate after her passing on January 22, 2018.

603 "UO Remembers Ursula Le Guin, Famed Author and Campus Icon," Around the O – University of Oregon Campus News, January 25, 2018, https://around.uoregon.edu/content/uo-remembers-ursula-le-guin-famed-author-and-campus-icon.
604 Ursula K. Le Guin Papers, University of Oregon Libraries, Special Collections and University Archives, http://archiveswest.orbiscascade.org/ark:/80444/xv926878#overview.
605 "Content Description," Ursula K. Le Guin Papers.
606 Ursula K. Le Guin Website and Social Media, Archive-It, https://www.archive-it.org/collections/10533.

Bibliography

#MeToo Web Archives Collection, Archive-It, https://archive-it.org/collections/10866.

#MeToo Project. Schlesinger Library. https://www.schlesinger-metooproject-radcliffe.org/.

"A Road Map for Digital Forensic Research." *Proceedings of the Digital Forensic Research Conference.* Utica, NY, August 7-8, 2001. https://dfrws.org/sites/default/files/session-files/a_road_map_for_digital_forensic_research.pdf.

"A Short History of the Web." CERN. https://home.cern/science/computing/birth-web/short-history-web.

"About." Archive-It. https://archive-it.org/blog/learn-more/.

About – Mormon Missionary Diaries." Brigham Young University. https://lib.byu.edu/collections/mormon-missionary-diaries/about/.

"About the Internet Archive." Internet Archive. https://archive.org/about/.

Abreu, Amelia. "The Collection and the Cloud." *The New Inquiry*. March 9, 2015. https://thenewinquiry.com/the-collection-and-the-cloud/.

Acker, Amelia. "Radical Appraisal Practices and the Mobile Forensic Imaginary." *Archive Journal* (2015). http://www.archivejournal.net/essays/radical-appraisal-practices-and-the-mobile-forensic-imaginary/.

Acker, Amelia and Jed Brubaker. "Death, Memorialization, and Social Media: A Platform Perspective for Personal Archives." *Archivaria* 77 (May 2014): 1-23.

Angello, Aaron. "To Archive or Not to Archive: The Resistant Potential of Digital Poetry." *Text Matters* 5 (2015): 13-27.

"Appraisal Issues." The Paradigm Project Workbook on Digital Private Papers, 2008. http://www.paradigm.ac.uk/workbook/appraisal/appraisal-issues.html.

Arbesman, Samuel. "The Forgotten Software that Inspired Our Modern World." *BBC News.* July 23, 2019. https://www.bbc.com/future/article/20190722-the-apple-software-that-inspired-the-internet.

"ARC_IA, Internet Archive ARC File Format." Sustainability of Digital Formats: Planning for Library of Congress Collections. https://www.loc.gov/preservation/digital/formats/fdd/fdd000235.shtml.

Archive Team. "GeoCities." https://www.archiveteam.org/index.php?title=GeoCities.

Avins, Mimi. "UCLA Buys Sontag's Archive." *Los Angeles Times*. January 26, 2002. https://www.latimes.com/archives/la-xpm-2002-jan-26-mn-24914-story.html.

Babbie, Earl. *The Practice of Social Research, 14th Edition.* Boston, MA: Cengage Learning, 2014.

Barlow, John Perry. "A Declaration of the Independence of Cyberspace." Electronic Frontier Foundation. February 8, 1996. https://www.eff.org/cyberspace-independence.

Barrett, Creighton. "Respect Which Fonds? Personal Archives and Family Businesses in Nova Scotia." *Archivaria* 76 (2013): 75-92.

Barry, Rick. "Web Sites as Recordkeeping and 'Recordmaking' Systems." *Information Management Journal* 38 (November/December 2004): 26-33.

Beagrie, Neil. "Plenty of Room at the Bottom? Personal Digital Libraries and Collections." *D-Lib Magazine* 11, no. 6 (2005). http://www.dlib.org/dlib/june05/beagrie/06beagrie.html.

Becker, Devin and Collier Nogues. "Saving-Over, Over-Saving, and the Future Mess of Writers' Digital Archives: A Survey Report on the Personal Digital Archiving Practices of Emerging Writers." *American Archivist* 75, no. 2 (2012): 482-513.

Belovari, Susanne. "Historians and Web Archives." *Archivaria* 83 (Spring 2017): 59-79.

Berkhofer, Robert F. *A Behavioral Approach to Historical Analysis*. New York: Free Press, 1969.

Berner, Richard C. *Archival Theory and Practice in the United States: A Historical Analysis*. University of Washington Press, 1983.

Biber, Katherine and Trish Luker. "Evidence and the Archive: Ethics, Aesthetics, and Emotion." *Australian Feminist Law Journal* 40, no. 1 (2014): 1-13.

Bowker, Geoffrey C. and Susan Leigh Star. *Sorting Things Out: Classification and Its Consequences.* Cambridge: MIT Press, 1999.

boyd, danah m. and Nicole B. Ellison. "Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication* 13, no. 1 (2008): 210-230.

Brennan, Bonnie S. *Qualitative Research Methods for Media Studies*. New York: Routledge (2017).

Briet, Suzanne. *What is Documentation?* Translated by Ronald E. Day and Laurent Martinet. Lanham, MD: The Scarecrow Press, 2006.

Brown, Thomas Elton. "The Society of American Archivists Confronts the Computer." *American Archivist* 47, no. 4 (Fall 1984): 366-382.

Buckland, Michael K. "What is a 'Document?'" *Journal of the American Society for Information Science* 48, no. 9 (1997): 804-809.

Burrows, Toby. "Personal Electronic Archives: Collecting the Digital Me." *OCLC Systems and Services: International Digital Library Perspectives* 22, no. 2 (2006): 85-88.

Busha, Charles and Stephen P. Harter. *Research Methods in Librarianship: Techniques and Interpretations*. New York, NY: Academic Press, 1980.

Cappon, Lester J. "Historical Manuscripts as Archives: Some Definitions and Their Application." *The American Archivist* 19, no. 2 (1956): 101-110.

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

Casey, Carol. "The Cyberarchive: A Look at the Storage and Preservation of Web Sites." *College & Research Libraries 59*, no. 4 (1998): 304-310.

Cibangu, Sylvain K. "A Memo of Qualitative Research for Information Science: Toward Theory Construction." *Journal of Documentation* 69, no. 2 (2013): 194-213.

Cohen, Patricia. "Fending of Digital Decay, Bit by Bit." *The New York Times*. March 16, 2010. https://www.nytimes.com/2010/03/16/books/16archive.html.

Coles, Laura. "Organizing Archival Material." In *A Manual for Small Archives*. Vancouver: Archives Association of British Columbia, 1998.

"Collections Policy Statements and Overviews." Library of Congress. https://www.loc.gov/acq/devpol/.

Colon-Marrero, Elena and Allison Hughes. "Toni Morrison's Born-Digital Material." *Mudd Manuscript Library Blog.* August 26, 2015. https://blogs.princeton.edu/mudd/2015/08/toni-morrisons-born-digital-material/.

Colon-Marrero, Elena and Allison Hughes, "Toni Morrison Collection," BitCurator Consortium Case Studies, https://www.bitcuratorconsortium.org/case-studies/toni-morrison-collection.

"Contribute." #MeToo Collection. https://www.schlesinger-metooproject-radcliffe.org/contribute.

"Contribute to the #metoo Project at Schlesinger Library." #MeToo Collection Submission Form. https://harvard.az1.qualtrics.com/jfe/form/SV_73eGeqtEoHPSsh7.

Cook, Terry. "Archival Science and Postmodernism: New Formulations for Old Concepts." *Archival Science* 1 (2001): 3-24.

Cook, Terry. "Electronic Records, Paper Minds: The Revolution in Information Management and Archives in the Post-Custodial and Post-Modernist Era," *Archives & Manuscripts* 22, no. 2 (1994): 300-328.

Cook, Terry. "What is Past is Prologue: A History of Archival Ideas since 1898 and Future Paradigm Shift." *Archivaria* 43 (February 1997): 17-63.

Cook, Terry and Joan Schwartz. "Archives, Records, and Power: From (Postmodern) Theory to (Archival) Performance." *Archival Science* 2 (2002): 171-185. https://doi.org/10.1007/BF02435620.

Cook, Terry and Joan Schwartz. "Archives, Records, and Power: The Making of Modern Memory." *Archival Science* 2 (2002): 1-19.

Cox, Richard J. *Managing Records as Evidence and Information*. Westport, CT: Quorum, 2001.

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

Cunningham, Adrian. "Beyond the Pale? The 'Flinty' Relationship Between Archivists Who Collect the Private Records of Individuals and the Rest of the Profession," *Archives and Manuscripts* 24, no. 1 (1996): 20-26.

Cunningham, Adrian. "The Archival Management of Personal Records in Electronic Form: Some Suggestions." *Archives & Manuscripts* 22, no. 1 (1994): 94-105.

Cunningham, Adrian. "Waiting for the Ghost Train." *Archival Issues* 24, no. 1 (1999): 55-64.

Cushing, Amber L. "Highlighting the Archives Perspective in the Personal Digital Archiving Discussion." *Library Hi Tech* 28, no. 2 (2010): 301-312.

Danielson, Elena S. *The Ethical Archivist.* Chicago: Society of American Archivists, 2010.

"Data Storage Devices." Media Archaeology Lab. https://mediaarchaeologylab.com/catalogue/peripheral/data-storage-devices/.

Davis, Allison P. "Susan Sontag, Sephora Shopper." *The* Cut. October 27, 2014, https://www.thecut.com/2014/10/susan-sontag-sephora-shopper.html.

"dBASE Table File Format (DBF)." Sustainability of Digital Formats: Planning for Library of Congress Collections. https://www.loc.gov/preservation/digital/formats/fdd/fdd000325.shtml.

Documenting the Now. https://www.docnow.io/.

Douglas, Jennifer. "Getting Personal: Personal Archives in Archival Programs and Curricula." *Education for Information* 33, no. 2 (2017): 89-105.

Douglas, Jennifer and Allison Mills. "From the Sidelines to the Center: Reconsidering the Potential of the Personal in Archives. *Archival Science* 18, no. 3 (2018): 257-277.

Drake, Jarrett. "The University Archives and its Focus on Fixity." Mudd Manuscript Library Blog. October 10, 2014. https://blogs.princeton.edu/mudd/2014/10/the-university-archives-and-its-focus-on-fixity/.

Durdella, Nathan. *Qualitative Dissertation Methodology*. London: Sage Publications, 2018.

Duranti, Luciana and Kenneth Thibodeau. "The Concept of the Record in Interactive, Experiential and Dynamic Environments: The View of InterPARES." *Archival Science* 6, no. 1 (2006): 13-68.

Dutton, William H., Everett M. Rogers, and Suk-Ho Jun. "Diffusion and Social Impacts of Personal Computers." *Communication Resear*ch 14, no. 2 (April 1987): 219-250.

Dwoskin, Elizabeth. "In a Single Tweet, as Man Pieces of Metadata as There Are Characters." Wall Street Journal. June 6, 2014.

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

"Emory Acquires Rushdie Archive." *Georgia Library Quarterly* 43, no. 4 (2007): 22.

Ernst, Wolfgang. *Digital Memory and the Archive*. Minneapolis: University of Minnesota Press. 2013.

Erway, Ricky. "You've Got to Walk Before You Can Run: First Steps for Managing Born-Digital Content Received on Physical Media." *OCLC Research*. August 2012. https://www.oclc.org/content/dam/research/publications/library/2012/2012-06.pdf.

Ethics Statement," #MeToo Collection, https://www.schlesinger-metooproject-radcliffe.org/ethics-statement.

Espenschied, Dragan. "README." GeoCities. https://github.com/despens/Geocities/blob/master/README.

Farr, Erika and Dorothy Waugh. “Salman Rushdie Archive.” BitCurator Consortium. https://www.bitcuratorconsortium.org/case-study/salman-rushdie-archive.

“FC5025.” Device Side Data. http://www.deviceside.com/fc5025.html.

File, Thom. “Computer and Internet Use in the United States.” *U.S. Census Bureau*. May 2013. https://www.census.gov/prod/2013pubs/p20-569.pdf.

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

“Fixity and Checksums.” Digital Preservation Coalition. https://www.dpconline.org/handbook/technical-solutions-and-tools/fixity-and-checksums.

Fleming, Barbara Morgan, Sandra Riegle, and Wesley Fryer. “Narrative Inquiry in Archival Work.” In *Narrative Inquiry: Experience and Story in Qualitative Research.* San Francisco: Jossey Bass Publishers (2000): 81-98.

“For Site Owners.” Web Archiving at the Library of Congress. https://www.loc.gov/programs/web-archiving/for-site-owners/.

Forstrom, Michael. “Managing Electronic Records in Manuscript Collections: A Case Study from the Beinecke Rare Book and Manuscript Library.” *American Archivist* 72, no. 2 (2009): 460-477.

“FRED.” Digital Intelligence. https://digitalintelligence.com/products/fred/.

Gabrys, Jennifer. *Digital Rubbish: A Natural History of Electronics*. Ann Arbor: University of Michigan Press, 2013.

Gardner, Philip. “Historical Analysis.” In *SAGE Dictionary of Social Research Methods*. London: SAGE Publications, 2006.

Garret, Jeffrey. “An Evaluation of Web Archiving Programs in the U.S. Relevant to International and Area Studies.” Center for Research Libraries Report, February 2019, https://www.crl.edu/sites/default/files/reports/Garrett%20Report%202019.pdf.

“GeoCities – The PATCHED Torrent.” The Pirate Bay, uploaded April 29, 2011. https://thepiratebay.org/description.php?id=6353395.

Georas, Chloé S. “Networked Memory Project: A Policy Thought Experiment for the Archiving of Social Networks by the Library of Congress.” *Laws* 3 (2014): 469-508.

Gilliland-Swetland, Anne. “Enduring Paradigm, New Opportunities: The Value of the Archival Perspective in the Digital Environment.” *CLIR Report*. February 2000. https://www.clir.org/pubs/reports/pub89/.

Gilliland-Swetland, Luke J. “The Provenance of a Profession: The Permanence of the Public Archives and Historical Manuscripts Traditions in American Archival History.” *The American Archivist* 54, no. 2 (1991): 160-175.

Gorman, G. E. and Peter Robert Clayton. *Qualitative Research for the Information Professional: A Practical Handbook.* Library Association Publishing, 1997.

Granger, Steward. “Emulation as a Digital Preservation Strategy.” *D-Lib Magazine* 6, no. 10 (2000): http://www.dlib.org/dlib/october00/granger/10granger.html.

Grotke, Abbie. “Legal Issues in Web Archiving.” *The Signal*. May 30, 2012. https://blogs.loc.gov/thesignal/2012/05/legal-issues-in-web-archiving/.

Hale, Scott A., Grant Blake, and Victoria D. Alexander. “Live versus Archive: Comparing a Web Archive to a Population of Web Pages.” In *The Web as History.* Edited by Niels Brügger and Ralph Schroeder. London: UCL Press, 2017.

Harris, Verne. “On the Back of a Tiger: Deconstructive Possibilities in ‘Evidence of Me.’” *Archives and Manuscripts* 29, no. 1 (2001): 8-21.

Hedstrom, Margaret. “Archives and Collective Memory: More than a Metaphor, Less than an Analogy.” In *Currents of Archival Thinking*, 2nd Edition. Edited by Heather MacNeil and Terry Eastwood. Santa Barbara, CA: Libraries Unlimited, 2017.

Hedstrom, Margaret. “Computers, Privacy, and Research Access to Confidential Information.” *Midwestern Archivist* 6, no. 1 (1981): 5-18.

Hedstrom, Margaret. and Christopher A. Lee. “Significant Properties of Digital Objects: Definitions, Applications, Implications.” *Proceedings of the DLM Forum* 2002. https://ils.unc.edu/callee/sigprops_dlm2002.pdf.

Henniger, Maureen and Paul Scifleet. “How Are the New Documents of Social Networks Shaping Our Cultural Memory?” *Journal of Documentation* 72, no. 2 (February 2016): 277-298. https://doi.org/10.1108/JD-06-2015-0069.

“History and Vision.” Me Too. https://metoomvmt.org/about/#history.

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

Hobbs, Catherine. “Reenvisioning the Personal: Reframing Traces of Individual Life.” In *Currents of Archival Thinking* 2nd ed. Santa Barbara: Libraries Unlimited (2010): 213-241.

Hodson, Sara. “Archives on the Web: Unlocking Collections While Safeguarding Privacy.” *First Monday* 11, no. 8 (2006): https://firstmonday.org/ojs/index.php/fm/article/view/1389.

Hodson, Sara S. “Private Lives: Confidentiality in Manuscript Collections.” *Rare Books and Manuscript Librarianship* 6, no. 2 (1991): 108-118.

“Home.” Katie Lee. https://www.katydoodit.com/.

Hou, Heng. “Concept, Process, and Principle: Research on the Privacy Protection in Web Archiving.” *Proceedings of the Doctoral Consortium of the 18th International Conference on Asia-Pacific Digital Libraries.* http://icadl2016.org/apis/14_Hou.pdf.

Houghton, Bernadette. “Preservation Challenges in the Digital Age.” *D-Lib Magazine* 22, no. 7 (2016): http://www.dlib.org/dlib/july16/houghton/07houghton.html.

Hu, Tung-Hui. *A Prehistory of the Cloud.* Cambridge, MA: The MIT Press, 2016.

Hunter, Gregory S. *Developing and Maintaining Practical Archives*. New York: Neil Schumann, 2004.

“Information Management: Challenges in Managing and Preserving Electronic Records: GAO-02-586.” United States Government Accounting Office. 2002, 45. https://www.gao.gov/products/GAO-02-586.

International Council on Archives. “Code of Ethics.” https://www.ica.org/sites/default/files/ICA_1996-09-06_code%20of%20ethics_EN.pdf.

Jackson, Stephen J. “Rethinking Repair.” In *Media Technologies: Essays on Communication, Materiality, and Society*. MIT Press (2014).

Jenkinson, Sir Hilary. *A Manual of Archive Administration Including the Problem of War Archives and Archive Making*. London, Oxford, New York: The Clarendon Press, 1922.

Jenkinson, Sir Hilary. *The English Archivist: A New Profession*. An inaugural lecture for a new course in archive administration delivered at University College, London. October 14, 1947.

Josias, Anthea. “Toward an Understanding of Archives as a Feature of Collective Memory.” *Archival Science* 11, no. 1-2 (March 2011): 95-112.

Jules, Bergis. “Some Thoughts on Ethics and DocNow.” *DocNow News*. June 3, 2016. https://news.docnow.io/some-thoughts-on-ethics-and-docnow-d19cfec427f2.

Kahle, Brewster. “Preserving the Internet.” *Scientific American* 276, no. 3 (1997): 82-83.

Kahle, Brewster. “Universal Access to All Knowledge.” *The Long Now Foundation*. November 30, 2011. http://longnow.org/seminars/02011/nov/30/universal-access-all-knowledge/.

Kaplan, Jeff. “The Fourth Generation Petabox.” *Internet Archive Blogs*. July 27, 2010. https://blog.archive.org/2010/07/27/the-fourth-generation-petabox/.

Katie Lee Collection, 1719-2019. Northern Arizona University. http://www.azarchivesonline.org/xtf/view?docId=ead/nau/lee_katie.xml&doc.view=print;chunk.id=0.

Kaufman, Lauren. “To Opt-In or Opt-Out?” *Popular Privacy*. March 6, 2020. https://medium.com/popular-privacy/to-opt-in-or-opt-out-5f14a10bae24.

Kelly, Mat, Michele C. Weigle. “WARCreate: Create Wayback-Consumable WARC Files from Any Webpage.” *JCDL ‘12: Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries* (June 2012): 437-438.

Kenney, Anne and Nancy McGovern. “Preservation Risk Management for Web Resources.” *Information Management Journal* 36, no. 5 (2002): 52-61.

Ketelaar, Eric. “Tacit Narratives: The Meaning of Archives.” *Archival Science* 12, no. 2 (2012): 131-141.

Khomami, Nadia. “#MeToo: How a Hashtag Became a Rallying Cry Against Sexual Harassment.” *The Guardian*. October 13, 2017. https://www.theguardian.com/world/2017/oct/20/women-worldwide-use-hashtag-metoo-against-sexual-harassment.

Kim, Julia Y. “Casting a Shadow, Coming out of the Dark: Digital Forensics with Personal Legacy Media.” Master of Arts Thesis. Moving Image Archiving and Preservation Program. New York University. September 2014.

Kirschenbaum, Matthew. “About the Deena Larsen Collection.” 2009. https://archive.mith.umd.edu/larsen/about/about/index.html.

Kirschenbaum, Matthew. “Software, It’s a Thing.” *Matthew Kirschenbaum*. July 25, 2014. https://medium.com/@mkirschenbaum/software-its-a-thing-a550448d0ed3.

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

Kirschenbaum, Matthew. *Track Changes: A Literary History of Word Processing*. Cambridge, MA: The Belknap Press of Harvard University, 2016.

Kirschenbaum, Matthew, Erika Farr, Kari Kraus, Naomi Nelson, Catherine Stollar Peters, Gabriela Redwine, Doug Reside. "Approaches to Managing and Collecting Born-Digital Literary Materials for Scholarly Use." White paper. 2009. https://drum.lib.umd.edu/handle/1903/9787.

Knight, Gareth. "The Forensic Curator: Digital Forensics as a Solution to Addressing the Curatorial Challenges Posed by Personal Digital Archives." *International Journal of Digital Curation* 7, no. 2 (2012): 40-63.

Kohler Riessman, Catherine. "Narrative Analysis." In *SAGE Dictionary of Social Research Methods*, ed. Victor Jupp. London: Sage Publications, 2006.

Kuny, Terry. "A Digital Dark Ages? Challenges in the Preservation of Electronic Information." *International Preservation News* 17 (May 1998): https://www.ifla.org/files/assets/pac/ipn/17-98.pdf.

Langdon, John. "Describing the Digital: The Archival Cataloguing of Born-Digital Personal Papers." *Archives and Records* 37, no. 1 (2016): 37-52.

Larsen, Deena. "Artist's Statement." https://archive.mith.umd.edu/larsen/items/show/165/index.html.

Lee, Christopher A. "Archival Application of Digital Forensics Methods for Authenticity, Description, and Access Provision." *Comma* 2 (2012): 133-140.

Lee, Christopher A. "Introduction." In *I, Digital: Personal Collections in the Digital Era.* Ed. Christopher A. Lee. Chicago: Society of American Archivists, 2011.

Leggit, Bob. "The History of Twitter in Profile Pages: 2006 to 2015." Twirpz, June 3, 2015. https://twirpz.wordpress.com/2015/06/03/the-history-of-twitter-in-profile-pages-2006-to-2015/.

Littman, Justin. "Web archiving and/or/vs Social Media API Archiving." Social Feed Manager. December 12, 2017. https://gwu-libraries.github.io/sfm-ui/posts/2017-12-13-web-social-media-archiving.

Littman, Justin, Daniel Chudnov, Daniel Kerchner, Christie Peterson, Yecheng Tan, Rachel Trent, Rajat Vij, & Laura Wrubel, "API-Based Social Media Archiving as a Form of Web Archiving," *International Journal on Digital Libraries* 19 (2018): 21-38.

Lowry, Charles B. "Putting the Pieces Together – Essential Technologies for the Virtual Library." *Journal of Academic Librarianship* 21 (July 1995): 297-300.

Lukesh, Susan S. “E-mail and Potential Loss to the Future Archives and Scholarship: The Dog That Didn’t Bark.” *First Monday* 4, no. 9 (1999): https://journals.uic.edu/ojs/index.php/fm/article/view/692.

Lyons, Bertram. “There Will Be No Digital Dark Age.” *Issues & Advocacy*. May 11, 2016. https://issuesandadvocacy.wordpress.com/2016/05/11/there-will-be-no-digital-dark-age/.

Maffeo, Lauren. “Google’s Vint Cerf on How to Prevent a Digital Dark Age.” *The Guardian*. May 29, 2015. https://www.theguardian.com/media-network/2015/may/29/googles-vint-cerf-prevent-digital-dark-age.

Mander Jones, Phyllis. “Archivists and the Association.” *The Australian Library Journal* 1, no. 4 (April 1952).

Marshall, Catherine C. “Rethinking Personal Digital Archiving Part 1: Four Challenges from the Field.” *D-Lib Magazine* 14, no. 3 (2008): http://www.dlib.org/dlib/march08/marshall/03marshall-pt1.html.

Marshall, Catherine C., Frank McCown, and Michael L. Nelson, “Evaluating Personal Archiving Strategies for Internet-Based Information,” Proceedings of Archiving 2007 Arlington, VA, (May 21-24, 2007): 151-156.

Marshall, Catherine C. and Frank Shipman. “Who Owns the Social Web?” *Communications of the ACM* 60 (2017): 52-61.

Mason, Richard O. “Four Ethical Issues of the Information Age.” *MIS Quarterly* 10, no. 1 (1986): 5-14.

McCree, Mary Lynn. “Good Sense and Good Judgment: Defining Collections and Collecting.” In *A Modern Archives Reader: Basic Readings on Theory and Practice*. Edited by Maygene F. Daniels and Timothy Walsh. Washington DC: National Archives Trust Board, 1984.

McDonald, Robert A.J. “Acquiring and Preserving Private Records: Cultural Versus Administrative Perspectives,” *Archivaria* 38 (January 1994): 162-163.

McGill, Andrew. “Can Twitter Fit Inside the Library of Congress?” *The Atlantic*. August 4, 2016. https://www.theatlantic.com/technology/archive/2016/08/can-twitter-fit-inside-the-library-of-congress/494339/.

McKemmish, Sue. “Evidence of Me.” *The Australian Library Journal* 45, no. 3 (January 1996): 174-187.

McKemmish, “Evidence of Me… In a Digital World,” in *I, Digital: Personal Collections in the Digital Era*, edited by Christopher A. Lee. Chicago: Society of American Archivists, 2011: 115-148.

McKemmish, Sue and Michael Piggott. “Toward the Archival Multiverse: Challenging the Binary Opposition of the Personal and Corporate Archive in Modern Archival Theory and Practice.” *Archivaria* 76 (2013): 111-114.

McKenzie, Pamela J. and Elisabeth Davies. “Genre Systems and ‘Keeping Track’ in Everyday Life.” *Archival Science* 12, no. 4. (2012): 437-460.

McLemee, Scott. “It Seemed Like a Good Idea at the Time.” *Inside Higher Ed*. January 5, 2018, https://www.insidehighered.com/views/2018/01/05/essay-library-congresss-twitter-archive.

Media Archaeology Lab, https://mediaarchaeologylab.com/.

Mell, Peter and Timothy Grance. The NIST Definition of Cloud Computing, Special Publication 800-145. National Institute of Standards and Technology. https://www.profsandhu.com/cs6393_s13/nist-SP800-145.pdf.

Memory Lab Nework, https://memorylabnetwork.github.io/.

Milligan, Ian. “Welcome to the Web: The Online Communities of GeoCities During the Early Years of the World Wide Web.” In *The Web as History.* Edited by Niels Brügger and Ralph Schroeder. London: UCL Press, 2017.

Mitchell, William J. *Placing Words: Symbols, Space, and the City.* Cambridge, MA: The MIT Press, 2005.

Montfort, Nick. “Continuous Paper: The Early Materiality and Workings of Electronic Literature.” *MLA Convention*. December 28, 2004, https://nickm.com/writing/essays/continuous_paper_mla.html.

Monua, Janah. “GeoCities Copes with Rapid Growth.” *InformationWeek*. December 7, 1998.

Mormon Missionary Collections. Brigham Young University. https://lib.byu.edu/collections/mormon-missionary-collections/.

Morris-Chott, Brooke. “Archiving Controversial Digital Materials,” Radcliffe Institute for Advanced Study News, August 8, 2019, https://www.radcliffe.harvard.edu/news/in-news/archiving-controversial-digital-materials.

Moser, Benjamin. “In the Sontag Archives.” *The New Yorker*. January 30, 2014. https://www.newyorker.com/books/page-turner/in-the-sontag-archives.

Moss, Michael and Tim Gollins. “Our Digital Legacy: An Archival Perspective.” *Journal of Contemporary Archival Studies* 4, no. 2 (2017): 1-29.

Muller, Samuel, J.A. Feith, and R. Fruin. *Manual for the Arrangement and Description of Archives*. Translation of the Second Edition by Arthur H. Leavitt. Chicago: Society of American Archivists, 2003.

"Muse." Mobisocial Lab at Stanford University. https://mobisocial.stanford.edu/muse/

Newburger, Eric C. "Home Computers and Internet Use in the United States: August 2000." *U.S. Census Bureau*. September 2001. https://www.census.gov/prod/2001pubs/p23-207.pdf.

Nissenbaum, Helen. "Privacy as Contextual Integrity." *Washington Law Review* 79, no. 1 (February 2004): 101-139.

Nu, Jinfang. "An Overview of Web Archiving." *D-Lib Magazine* 18, no. 3-4 (2012): http://www.dlib.org/dlib/march12/niu/03niu1.html.

O'Sullivan, Catherine. "Diaries, On-Line Diaries, and the Future Loss to Archives: Or, Blogs and the Blogging Bloggers Who Blog Them." *The American Archivist* 68, no. 1 (2005): 53-73.

Oliver, Gillian and Wendy M. Duff. "Genre Studies and Archives: Introduction to the Special Issue." *Archival Science* 12, no. 4 (2012): 373-376.

Oliver, Gillian, Yunhyong Kim, and Seamus Ross. "Documentary Genre and Digital Recordkeeping: Red Herring or a Way Forward?" *Archival Science* 8, no. 4 (December 1, 2008): 295-305.

One Terabyte of Kilobyte Age. https://blog.geocities.institute/about.

Osterberg, Gayle. "Update on the Twitter Archive at the Library of Congress." *Library of Congress Blog.* January 4, 2013. https://blogs.loc.gov/loc/2013/01/update-on-the-twitter-archive-at-the-library-of-congress/.

Osterberg, Gayle. "Update on the Twitter Archive at the Library of Congress." *Library of Congress Blog*. December 26, 2017. https://blogs.loc.gov/loc/2017/12/update-on-the-twitter-archive-at-the-library-of-congress-2/?loclr=blogloc.

Owens, Trevor. "Glitching Files for Understanding: Avoiding Screen Essentialism in Three Easy Steps." *The Signal*. November 5, 2012. https://blogs.loc.gov/thesignal/2012/11/glitching-files-for-understanding-avoiding-screen-essentialism-in-three-easy-steps/.

Pace, Andrew. "Coming Full Circle: Digital Preservation: Everything New is Old Again." *Computers in Libraries* 20, no. 2 (2000): 55-59.

Pan, Yin, Bill Stackpole, and Luther Troell. "Computer Forensics Technologies for Personally Identifiable Information Detection and Audits." *ISACA* 2 (2010): 1-6.

Pearce-Moses, Richard. “The Perfect and the Possible: Becoming a Digital Archivist.” Paper presented at the Conference of Inter-Mountain Archivists. 1May 12, 2006. www.lib.az.us/about/annualreports/2006/the.perfect.and.the.possible.becoming.a.digital.archivist.pdf.

Pollard, Riva. “The Appraisal of Personal Papers: A Critical Literature Review.” *Archivaria* 52 (February 2001): 136-150.

Powell, Graeme. “The Collecting of Personal and Private Papers in Australia.” *Archives & Manuscripts* 24, no. 1 (1996): 62-77.

Praetzellis, Maria. “Glossary of Archive-It and Web Archiving Terms.” https://support.archive-it.org/hc/en-us/articles/208111686-Glossary-of-Archive-It-and-Web-Archiving-Terms.

Rao, “Yahoo Quietly Pulls the Plug on GeoCities.” *TechCrunch*. April 23, 2009. https://techcrunch.com/2009/04/23/yahoo-quietly-pulls-the-plug-on-geocities/.

Raymond, Matt. “The Library and Twitter: An FAQ.” *Library and Congress Blog*. April 28, 2010. http://blogs.loc.gov/loc/2010/04/the-library-and-twitter-an-faq/.

Redwine, Gabriela. “Personal Digital Archiving.” Digital Preservation Coalition Technology Watch Report 15-01. December 2015. https://www.dpconline.org/docs/technology-watch-reports/1460-twr15-01/file.

Reyes, Vanessa. “We Created It, Now How Do We Save It? Issues in Preserving Personal Information, A Review.” *Preservation, Digital Technology & Culture* 42, no. 3 (2013): 150-154.

Robertson, Tara. “Not All Information Wants to Be Free.” LITA Forum Keynote 2016. https://tararobertson.ca/2016/lita-keynote/.

Rockmore, Dan. “The Digital Life of Salman Rushdie.” *The New Yorker*. July 29, 2014. https://www.newyorker.com/tech/annals-of-technology/digital-life-salman-rushdie.

Rogers, Corinne and Luciana Duranti. “Ethics in the Cloud.” *Journal of Contemporary Archival Studies* 4, no. 2 (2017): 1-11.

Rothenberg, Jeff. “Ensuring the Longevity of Digital Information.” Council on Library and Information Resources, Report, February 22, 1999. https://clir.wordpress.clir.org/wp-content/uploads/sites/6/ensuring.pdf.

Rothstein, Edward. “Displaying the Dinosaurs of the Digital Age.” *New York Times*. September 28, 2012. https://www.nytimes.com/2012/09/29/arts/design/computer-history-museum-in-mountain-view-calif.html.

Roy, Hugo. "ToS;DR in Action: 'I Have Read and Agree to the Terms." *Terms of Service; Didn't Read Blog.* June 10, 2013. https://tosdr.org/blog/tosdr-in-action-i-have-read-and-agree.html.

Roy, Hugo. "'What is Yours Stays Yours': Oh, Really?" *Terms of Service; Didn't Read Blog.* June 17, 2013. https://tosdr.org/blog/tosdr-in-action-owned.html.

Rule, James B. *Privacy in Peril*. New York: Oxford University Press, 2007.

Salman Rushdie Papers, Emory University Manuscripts, Archives, and Rare Books Library. https://findingaids.library.emory.edu/documents/rushdie1000/.

"Save Unique, At-Risk Digital Content." OCLC Member Stories. https://www.oclc.org/en/member-stories/ucla.html.

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

"Schlesinger Library Awarded Grant to Create Comprehensive Digital Media Archive of #metoo." Radcliffe Institute for Advanced Study Harvard University. June 25, 2018. https://www.radcliffe.harvard.edu/news/press-releases/schlesinger-library-awarded-grant-digital-archive-metoo.

Schmidt, Jeremy and Jacquelyn Ardam, "On Excess: Sontag's Born-Digital Archive." *LA Review of Books*. October 26, 2014. https://lareviewofbooks.org/article/excess-susan-sontags-born-digital-archive/.

Scott, Jason. "The Splendiferous Story of Archive Team." Personal Digital Archiving Conference 2011. http://ascii.textfiles.com/archives/3029.

Sharma, Seemantani, "'How Tweet it is!': Have Twitter Archives Been Left in the Dark?" *Journal of Law, Technology, and Policy* no. 1 (2019): 49-78.

Shein, Cyndi. "From Accession to Access: A Born-Digital Materials Case Study." *Journal of Western Archives* 5, no. 1 (2014): 1-42.

Sinn, Donghee and Sue Yeon Syn. "Personal Documentation on a Social Network Site: Facebook, a Collection of Moments from Your Life?" *Archival Science* 14, no. 2 (2014):95-124.

Slade, Giles. *Made to Break: Technology and Obsolescence in America.* Cambridge, MA: Harvard University Press, 2006.

Smith, Abby. "Introduction." In *Enduring Paradigm, New Opportunities: The Value of the Archival Perspective in the Digital Environment.* Anne J. Gilliland-Swetland. Council on

Library and Information Resources Report 2000. https://www.clir.org/pubs/reports/pub89/.

Smith Rumsey, Abby. *When We Are No More: How Digital Memory is Shaping Our Future.* Bloomsbury Press, 2016.

Society of American Archivists. "A Guide to Deeds of Gift." https://www2.archivists.org/publications/brochures/deeds-of-gift#.V0kj984sj6U.

Society of American Archivists. "SAA Core Values Statement and Code of Ethics." https://www2.archivists.org/statements/saa-core-values-statement-and-code-of-ethics.

Society of American Archivists. *Dictionary of Archives Terminology.* https://dictionary.archivists.org/index.html.

Socio-Technical Sustainability Roadmap. https://sites.haa.pitt.edu/sustainabilityroadmap/.

Spragge, Shirley. "The Abdication Crisis: Are Archivists Giving Up their Cultural Responsibility?" *Archivaria* 40 (1995): 173-181.

Star, Susan Leigh. "The Ethnography of Infrastructure." *American Behavioral Scientist* 43, no. 3 (1999): 335-365.

Stone, Biz. "Tweet Preservation." Twitter Blog. April 14, 2010. https://blog.twitter.com/official/en_us/a/2010/tweet-preservation.html.

Strauss, Amanda. "#MeToo: A Glimpse into the Digital Vault." Schlesinger Newsletter, Fall 2018. https://www.radcliffe.harvard.edu/news/schlesinger-newsletter/metoo-glimpse-digital-vault.

Summers, Ed. "A Twitter ToS Deep Dive with Justin Littman." *DocNow News.* March 7, 2019. https://news.docnow.io/a-twitter-tos-deep-dive-with-justin-littman-728feb54c0fd.

Sumner, Maggie. "Qualitative Research." In *The SAGE Dictionary of Research Methods*. Edited by Victor Jupp. London: SAGE Publications, 2006.

Susan Sontag Papers. LSC.0612. UCLA Library Special Collections, Charles E. Young Research Library. https://oac.cdlib.org/findaid/ark:/13030/kt2489n7qw/.

Taylor, Nicholas. "The Average Lifespan of a Webpage." *The Signal.* November 8, 2011. https://blogs.loc.gov/thesignal/2011/11/the-average-lifespan-of-a-webpage/.

Thibodeau, Kenneth. "Overview of Technological Approaches to Digital Preservation and Challenges in the Coming Years. Council on Library and Information Resources. Report, 2002. https://www.clir.org/pubs/reports/pub107/thibodeau/.

Thomas S. Mullaney. Archive-It. https://archive-it.org/collections/10862.

Thomas S. Mullaney Papers SC1435SC1435. Stanford University Archives. http://pdf.oac.cdlib.org/pdf/stanford/uarc/sc1435.pdf.

Toni Morrison Papers, C1491. Manuscripts Division, Department of Special Collections, Princeton University Library. https://findingaids.princeton.edu/collections/C1491/c4181.

"Toni Morrison Papers to Reside at Princeton." Princeton University News. October 17, 2014. https://www.princeton.edu/news/2014/10/17/toni-morrison-papers-reside-princeton.

"Toni Morrison Papers Open for Research." *Princeton University Library*. June 8, 2016. https://library.princeton.edu/news/general/2016-06-08/toni-morrison-papers-open-research.

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

Twarc," Documenting the Now GitHub, https://github.com/DocNow/twarc.

United States Department of Commerce, Economics, and Statistics Administration. "The Growing Use of Computers." Bureau of the Census. April 1991. https://www2.census.gov/library/publications/1991/demographics/sb91-11.pdf.

University of North Carolina at Chapel Hill Confederate Monument Protests Collected Tweets, 2017-2020. https://finding-aids.lib.unc.edu/40486/#d1e88.

Update on the Twitter Archive at the Library of Congress. White Paper. December 2017. https://blogs.loc.gov/loc/files/2017/12/2017dec_twitter_white-paper.pdf.

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

Ursula K. Le Guin Website and Social Media. Archive-It. https://www.archive-it.org/collections/10533.

"Using the Wayback Machine." Internet Archive. https://help.archive.org/hc/en-us/articles/360004651732-Using-The-Wayback-Machine.

"Version 2." Previous Twitter Privacy Policy." https://twitter.com/en/privacy/previous/version_2.

"Version 3." Twitter Terms of Service. https://twitter.com/en/tos/previous/version_3.

"WARC, Web ARChive file format." Library of Congress. https://www.loc.gov/preservation/digital/formats/fdd/fdd000236.shtml.

Watters, Audrey. "How the Library of Congress is Building the Twitter Archive." *O'Reilly Radar*. June 2, 2011. http://radar.oreilly.com/2011/06/library-of-congress-twitter-archive.html.

Waugh, Dorothy and Elizabeth Russey Roke. "Second Generation Digital Archives: What We Learned from the Salman Rushdie Project." Personal Digital Archiving Conference, 2017. https://personaldigitalarchiving2017.sched.com/event/9JaJ/session-2-preserving-serving-pda-at-memory-institutions.

"Web Archiving." Library of Congress. https://www.loc.gov/programs/web-archiving/about-this-program/.

Weldon, Edward. "Archives and the Challenge of Change." *American Archivist* 46, no. 2 (Spring 1983): 125-134.

"What is a Web Server?" Mozilla Developer Network. https://developer.mozilla.org/en-US/docs/Learn/Common_questions/What_is_a_web_server.

"What is Web Archiving? History, Technology, Collections." *National Records of Scotland Open Book Blog*. November 24, 2017. https://blog.nrscotland.gov.uk/2017/11/24/what-is-web-archiving-history-technology-collections/.

"What We Collect." Stanford University Special Collections and University Archives. https://library.stanford.edu/spc/university-archives/about-archives/what-we-collect.

Williams, Caroline. "Personal Papers: Perceptions and Practices." In *What Are Archives? Cultural and Theoretical Perspectives: A Reader.* Edited by Louise Craven. London: Routledge, 2016.

Wilson, Jacqueline Z. and Frank Golding. "Latent Scrutiny: Personal Archives as Perpetual Mementos of the Official Gaze." *Archival Science* 16, no. 1 (2016): 93-109.

Wittenberg, Jamie and Celia Emmelhainz, "Access, Annotate, Export." In *The Complete Guide to Personal Digital Archiving*. Edited by Brianna Marshall. Chicago: ALA Editions, 2017.

"Yahoo Sets the Date of GeoCities' Death," *PC Mag*, July 10, 2009, https://www.pcmag.com/archive/yahoo-sets-the-date-of-geocities-death-242171.

Zimmer, Michael. "Is it Ethical to Harvest Public Twitter Accounts without Consent?" *Michael Zimmer*. February 12, 2010. https://www.michaelzimmer.org/2010/02/12/is-it-ethical-to-harvest-public-twitter-accounts-without-consent/.

Zimmer, Michael. “The Twitter Archive at the Library of Congress: Challenges for Information Practice and Information Policy.” *First Monday* 20, no. 7 (2015): https://firstmonday.org/article/view/5619/4653.

Zine Web Collection. Library of Congress. https://www.loc.gov/collections/zine-web-archive/about-this-collection/.