Casting a Shadow, Coming out of the Dark:

Digital Forensics with Personal Digital Legacy Media.

.

by
Julia Y. Kim

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Arts
Moving Image Archiving and Preservation Program
Department of Cinema Studies
New York University
September 2014
Advisor: Mona Jimenez

In the context of cultural archives, it is commonplace for there to be a backlog of born-digital hardware. The open secret of legacy digital backlog is something that digital forensics addresses by blending computer science with familiar archival principles such as "chain-of-custody" and "provenance." As more archives grapple with these digital holdings, however, the interdependencies and complexities of hardware, software, file formats, and intellectual property rights complicate preservation workflows. Best practices and standards have not yet been well developed or understood. A typical way to handle acquired hard-drives, for example, may be on a case-by-case basis, dependent upon the collection's characteristics, priorities, and access plans. However, such ad hoc workflows are time and resource intensive.

In recent years, conceptual frameworks and tools adopted from the field of "digital forensics" have made significant inroads within cultural archives and institutional practices. The field of digital forensics is increasingly seen as a solution to the backlog of obsolete media. But factors such as budgets, technical training, administrative workflow restructuring, and the still-obscure issue of researcher access continue to hinder many institutions from integrating digital forensics. Unfortunately, the dubious image of a hard drive archived in an acid-free file folder, inaccessible to the public and without any immediate plans for future access, is all too common. Many collecting institutions cannot meaningfully preserve obsolete digital material. Through the course of this work, I will highlight successful examples of digital forensics implementation in archival institutions.

In this thesis, I discuss the historical intersection between forensics and archives; document several contemporary case studies; and present preliminary conclusions that focus on researcher access. The hybrid complex personal collections that serve as case studies here benefit scholar-

critics in the digital humanities, academic and non-academic researchers, and archivists alike, but digital forensics has applications for all complex born-digital media. This work will be situated within the relatively new and ongoing history of the integration of digital forensics in archival workflows, while also drawing on the intersection of computer science and digital humanities. Thus, my methodology will include traditional research, case studies, and analysis.

This topic goes beyond the technical to include philosophical and ethical issues regarding private and personal materials, the inevitable copyright and intellectual property complications, and the largely unaddressed issues of researcher needs and technical responsibilities. Access issues will be explored at length in my concluding chapter. This research will also serve to clarify the limitations and applications of digital forensics in an archival setting. In exploring digital forensics, I will set aside discussion of well-known and trusted born-digital media. Well-documented removable media used for transmission, for example, would not need digital forensic detective work. This thesis addresses media that requires more intensive intervention due to, for example, a lack of documentation and technical knowledge. Practices for handling these types of materials have typically remained scantily documented, if at all, for discovery.

Acknowledgments

Special thanks to my thesis advisor Professor Mona Jimenez; her support, guidance, and patience throughout the past several years has been invaluable. Many thanks also to Don Mennerich, Digital Archivist of New York University for his generosity. His experiences as a pioneer forensic archivist in the archival world were essential to guiding my thesis development. Thank you to my editor, C Reis.

Thanks to the MIAP faculty and staff: Howard Besser, Dan Streible, and Alicia Kubes; and to the MIAP graduating class of 2013 and 2014 for their support and encouragement. Additional thanks to the New York Public Library Manuscript staff, including Tal Nadan and Alison Rhonemous. Thanks also to Erika Farr, Porter Olsen, and Kam Woods.

Chapter 1: History and Essential Concepts

Within the archiving and preservation community, the complexity of the preservation of born-digital media is well-known.[1] These issues are compounded with legacy removable media carriers.[2] Despite decades of born-digital media acquisition, archives often have, at best, "ad hoc and iterative practices" developed through trial and error, specified to each collection's processing.[3] According to Susan Davis's 2006 survey, 47% of collecting repositories had already acquired or planned to acquire born-digital collections. Yet 76% of these institutions did not have a processing policy accounting for born-digital acquisitions. Over half of those that did have a plan in place noted that the policy was indistinguishable from their paper-based acquisition policy.[4] In the qualitative feedback, opinions on this statistic ranged: some felt there was no need for addressing the change in format over content categories. More recently, a 2012 Association of Research Libraries survey found little change from the Davis findings: 78% of those surveyed indicated that born-digital records are stored on legacy media "as is." These statistics are significant; they are indicative of a de facto ineffective standard for addressing preservation issues presented by complex media.[5] The survey results are further indicative of the large gap between practice and theory; the debates surrounding migration vs. emulation are irrelevant to many institutions as they struggle to simply maintain born-digital records. While digital preservation

---

1 Ricky Erway defined born digital as "items created and managed in digital form," in "Defining Born-Digital," *OCLC Online Computer Library Center, Inc.,* November 2010, www.oclc.org/content/dam/research/activities/hiddencollections/borndigital.pdf?urlm=161291.

2 I define legacy to denote non-contemporaneous removable media.

3 Martin Gengenbach, "The Way We Do It Here": Mapping Digital Forensics Workflows in Collecting Institutions (master's thesis, University of North Carolina at Chapel Hill, August 2012, http://digitalcurationexchange.org/system/files/gengenbach-forensic-workflows-2012.pdf.

4 Susan E. Davis, "Electronic Records Planning in 'Collecting' Repositories," *American Archivist* 71:7 (2007): 178-179, http://archivists.metapress.com/content/024q2020828t7332/

5 Naomi L. Nelson et al, Managing Born-Digital Special Collections and Archival Materials, SPEC Kit 329. (Washington, DC Association of Research Libraries, 2012), 35, file:///C:/Users/Eyi%20Kim/Downloads/2h4qbi%20(1).pdf.

practice has grown tremendously this past decade, many institutions are unable to make the jump to managing the legacy of born-digital materials. Scholarship might better focus on navigating barriers to workflow integration. Meaningful access, the last step in any processing plan, is irrelevant, as these statistics indicate no current practical workflows for born-digital processing, the essential precursor to access. In the last decade, especially in the past two to three years, digital forensics has grown from a virtual non-presence to a much talked about emerging specialty within the field of digital preservation. Despite the rapid shift from relative obscurity to the latest buzzword in the archiving community, there is a long way to go before digital forensics's significant solutions for processing complex born-digital media can consistently adopted and applied.

1.1 Digital Forensics Fields

Digital forensics has its genesis in law enforcement and computer science. While today it branches into ancillary fields such as network forensics, its beginnings date back to the inception of personal computing in the 1970s. An authoritative definition drafted from the first Digital Forensics Research Workshop meeting in 2001 is as follows:

> The use of scientifically derived and proven methods towards the preservation, collection, validation, identification, analysis, interpretation, documentation, and presentation of digital evidence derived for the purpose of facilitating or furthering the reconstruction of events found to be criminal.[6]

This definition emphasizes digital forensics′s necessary dependence upon the scientific method for the end purpose of criminal investigation. While this is a good starting point for understanding digital forensics, it is of limited application for archives. Archivists, for example, do not need to meet a legal threshold of proof; they are tasked with reconstructing while avoiding interpretation. In the past several years, the term's shifting usage during its encroachment into archival workflows have imbued it with other connotations.

The first acknowledged pioneers of digital forensics, Seamus Ross and Ann McGow, published a 1999 paper advocating the implementation of forensics methods for pulling data off of obsolete media. Their methods include magnetic force microscopy and cryptography. Magnetic force microscopy (MFM) registers patterns of magnetic fluctuations on media, allowing one to see images of layers of overlaid data inscriptions on low density media. While these methodologies have not been followed by the archival community, their paper opened up new possibilities for interdisciplinary work with forensics communities. McGow and others pushed

[6] Gary Palmer, "A Road Map for Digital Forensics Research: Report from the First Digital Forensic Research Workshop," *DFRWS Technical Report,* August, 2001, 16, http://www.dfrws.org/2001/dfrws-rm-final.pdf.

archivists to understand that meaningful preservation goes far beyond maintaining macro conditions such as temperature and humidity. They helped re-frame the issue as not just preservation of the media carrier, but as an understanding of media's materiality in the form of inscribed bits.[7]

For computer scientists and archivists alike, Simson Garfinkel is the single most influential figure in the fields of digital forensics. Garfinkel, a computer scientist affiliated with the U.S. Naval Postgraduate School in Monterey, has authored a number of touchstone publications and has also developed widely adopted software tools. These tools include Guymager, an imaging tool, and the Sleuth Kit, a software suite for forensic analysis. Some of his influence in the archiving community is evident in that both have been adopted into the BitCurator suite of tools. With the maturation of the personal computing market, Garfinkel was able to uniquely influence the development of the field during the so-called "Golden Age" of forensics (1997-2007). Through that period of operating system normalization and tool development, Garfinkel and others made many strides in procedures and tools that are still relevant in archives.[8]

Digital forensics has now pushed to the forefront of archival developments. Major innovations have been made. Many institutions have added digital forensics components to their departments. Just this past decade, a number of well-publicized studies and projects have pushed the development of documented workflows for archivists, building on work from the early 2000s by institutions including Stanford University, the Bodleian Library, and the British Library. Additionally, attention has been paid to a curriculum for the training of archivists, including

---

[7] Seamus Ross and Ann Gow, "Digital Archaeology: Rescuing Neglected and Damaged Data: A JISC/NPO Study within the Electronic Libraries (eLib) Programme on the Preservation of Electronic Materials," (Glasgow, Humanities Advance Technology and Information Institute: 1999), ukoln.ac.uk/services/elib/papers/supporting/pdf/p2.pdf.

[8] Simson L. Garfinkel, "Digital Forensics Research: the Next 10 years," *Digital Investigation* 7(2010), S64-S73, http://dfrws.org/2010/proceedings/2010-308.pdf,

workshops and certifications. The Society for American Archivists, the University of Virginia's Rare Book School course "Born Digital Materials: Theory and Practice," and the University of Maryland's Humanities Intensive Learning and Teaching "Born-Digital Forensics" are all short-term workshops directed towards practicing archivists. [9] What follows is only a brief summary of influential collaborative and multi-institutional projects, as well as scholarly works:

"**Digital Forensics and Born-Digital Content in Cultural Heritage Collections**" **(**2010) by Matthew Kirschenbaum, Richard Ovenden, and Gabriela Redwine.

A definitive overview and the best available introduction of the convergence between digital forensics and archives.[10]

"**AIMS Born-Digital Collections: An Inter-Institutional Model for Stewardship**" (2012). This paper is the result of a two0-year Carnegie Mellon-funded project addressing born-digital preservation challenges common to special collections at research universities. [11] The paper addresses major steps in processing, providing specific collection processing plans from AIMS partner institutions. The report primarily address stewardship of born digital materials, but it emphasizes that a "constant (in collections) was the presence of legacy collections and anomalies."

---

[9] Matthew Kirschenbaum and Naomi Nelson. Rare Book School: Preliminary Reading List, rarebookschool.org/reading/libraries/l95/. Society of American Archivists. Digital Archives Specialist Curriculum and Certificate Program, http://www2.archivists.org/prof-education/das. Humanities Intensive Learning and Teaching, University of Maryland, dhtraining.org/hilt/course/born-digital-forensics/.

[10] Matthew Kirschenbaum, Richard Ovenden, Gabriela Redwine, "Digital Forensics and Born-Digital Content in Cultural Heritage Collections," Council on Library and Information Sources (2010), http://www.clir.org/pubs/abstract/reports/pub149.

[11] "Born-Digital Archives," AIMS blog, http://born-digital-archives.blogspot.co.uk.

"**Digital Forensics and Preservation**" (2012) by Jeremy Leighton John.

John provides a very detailed, thoughtful exploration of digital forensics applications to personal digital archives, signaling a major shift in the field in exploring this direction. In concluding, John states "there are some deep challenges ahead…but the forensic perspective is undoubtedly among the most promising source of insights and solutions."[12]

"**From Bitstreams to Heritage: Putting Digital Forensics into Practice in Collecting Institutions**" (2013) by Christopher Lee, Kam Woods, Matthew Kirschenbaum, and Alexandra Chassonoff. This paper is the year-end, mid-project report of the BitCurator project, an Andrew W. Mellon Foundation funded, multi-institutional, two year project to integrate pre-existing digital forensics tools for archival workflow adoption. Through the use of their expert advisory panels, they specifically address the archival (versus law enforcement) needs for privacy redaction and tools that work well with legacy materials.[13] BitCurator also references a downloadable product; it is a packaged suite of already available open-source modular tools. The project is currently focusing on adoption among archives across the country and world, but they have recently won a two-year Carnegie Mellon grant to fund BitCurator "access" to develop metadata discovery

[12] Jeremy John Leighton, "Digital Forensics and Preservation, Digital Preservation Coalition, 2012, 8, DOI: http://dx.doi.org/10.7207/twr12-03.

[13] Cal Lee, Kam Woods, Matthew Kirschenbaum, and Alexandra Chassanof, "From Bitstreams to Heritage: Putting Digital Forensics into Practice in Collecting Institutions," BitCurator Project (2013).

tools for online discovery and even access of documents.[14] While the tools and associated workshops have already had tremendous impact in the field, the new push for discovery also at least equal potential to change born-digital access norms.

While there have been tremendous strides in the last several years, the time to move from basic research to much more widespread implementation of digital forensics practices is now. From this overview of literature and project work in the past several years, we can note that digital forensics has only recently begun to gain momentum as a nascent applied sub-field of digital preservation. Major advocacy for digital forensics use within the archival community dates back several years. But as of now, implementation is delayed until further studies are published. One of the aims of the present study is to clarify and aid in this endeavor. I do not and cannot give any definitive answers, but this thesis should serve as a good primer not only for past and present problems, but also for possible solutions and paths towards more accessible collections in the future.

What we talk about when we talk about digital forensics

The evocation of digital forensics in the archival world, then, emphasizes different values, needs, and preconceptions. While there are inherent and persistent qualities, "digital forensics" can and does signify different things. This is simply a by-product of the many disciplines that contribute to digital forensics applications.[15] I find it most useful to move beyond the standard definition cited earlier by proposing the following, three-part definition of digital forensics in terms

[14] "Bitcurator Access," Bitcurator, modified May 14, 2014, access.bitcurator.net/.

[15] Jeremy John Leighton, "Digital Forensics and Preservation, Digital Preservation Coalition, 2012, 7-10, DOI: http://dx.doi.org/10.7207/twr12-03.

of: 1. Broad frameworks; 2. Workflows and practices, 3. Specific tool-sets. Each of these categories is also important to structuring legal forensic practices. In the following section, I will highlight some of the adoptions and challenges to archival adoption using these categories. When archivists speak of "digital forensics," they are not using the formal inherited definition, but allude instead to the archival adoption of the aforementioned categories above.

Framework

Forensics proponents in archives have all remarked on the happy consonance between traditional, long-established tenets underlying the archival sciences and the framework of digital forensics: the frameworks nearly directly map onto each other with striking correspondences between the two.[16] While many different digital forensic investigation frameworks have been proposed over the years, at its heart the process of forensics can be distilled into three steps: acquire, analyze, and interpret.[17] Acquire the evidence, analyze it using software and hardware tools, and then interpret the evidence to ultimately understand its significance. These intuitive steps guide and organize a digital forensics investigation within a legal and criminal framework, but they also work within established archival frameworks.[18] The "acquisition" of a collection

[16] The Latin etymology of forensics- "of the court."
[17] Jeremy John Leighton, "Digital Forensics and Preservation, 8.-9.
[18] Gareth Knight, "The Forensic Curator: Digital Forensics as a Solution to Addressing the Curatorial Challenges Posed by Personal Digital Archives. The International Journal of Digital Curation 7:2, 2012, 44. Gareth Knight synthesizes much work into a 6 step framework: Prepare, Acquire, Examine, Analyze, Report, and Review. Many different frameworks have been proposed over the years.

or evidence for both fields is a carefully negotiated and controlled process. As part of the acquisition process, both seek to assess and maintain authenticity. In the legal forensic framework, this tenet is an essential underpinning, starting with the successful establishment of a crime scene with the use of write-blockers to ensure that evidence is not altered. This is similar to the archival establishment of provenance and an unbroken chain-of-custody.[19] "Chain-of-custody," a term specific in both archival sciences and legal, criminal digital forensics specifically verifies the controlled handling of the material throughout its successions, whether in a criminal seizure or in an archival accession.[20]

Law Enforcement Digital Forensics	**Archival Sciences**
Acquire without alteration, maintain and establish authenticity	Acquisition process without alteration.[21] (provenance, chain-of custody)
Analyze evidence in an accountable and repeated fashion	Analyze for privacy and security concerns
Interpret: Exploit private data for prosecution	Interpret: Intellectually arrange and make essential aspects accessible.

Figure 1: Similar Concepts

For both applications, however, these tenets are not black and white; they must be evaluated holistically. Authenticity, in the age of endless and virtually perfect duplication, is subject to evaluation. In fact, within the validation procedure, archival institutions may, by virtue of

[19] Chain of custody in the legal forensics context means "the chronological documentation or paper trail, showing the seizure, custody, control, transfer, analysis, and disposition of evidence, physical or electronic." http://www.uio.no/studier/emner/matnat/ifi/INF3510/v12/learningdocs/WS-INF3510-2012-L11-QA.pdf. Within an archival context according to the Society for American Archivists: n. ~ 1. Records · the succession of offices or persons who have held materials from the moment they were created. - 2. Law · the succession of officers or individuals who have held real evidence from the moment it is obtained until presented in court. http://www2.archivists.org/glossary/terms/c/chain-of-custody.

[20] Miriely Guerrero, "Removable Media and the Use of Digital Forensics," Bentley Historical Library, Digital Curation, July 2, 2012, 22, http://bentley.umich.edu/dchome/resources/RemovableMediaReport.pdf.

[21] "Code of Ethics," American Institute of Conservation, http://www.conservation-us.org/about-us/core-documents/code-of-ethics#.U_UFz_ldVOk.

accepting material, play a role in *bestowing* authenticity.[22] That is, as much as they ensure, they also play a role in signifying and sealing. Another way to approach this preeminent concern is through exploring "evidentiary value" in both spheres. As archivist Glenda Acland noted in "Archival Authenticity in a Digital Age," the "pivot of archival science is evidence, not information."[23] It is the first task of both fields to ensure the "ground truth," while the validity of the information is, at best, secondary.[24] Johns notes that the evidentiary interpretation is the outcome of a "multi-evidential approach" accounting for the entirety of the picture formed through the accumulation of "traces," or leftover signs of interactions. Events must be reconstructed, timelines assessed, and some sort of putative meaning derived. Complex digital objects have only forced archivists to reconfigure these principles in the face of endless and exact duplication. Working with abstractions, the narrative of events and actions must be reconstructed.[25]

Another cornerstone in the development of a legal forensic investigation is the evidentiary need for "accountable" and "repeatable" methods of analysis, resulting from the need to meet legal thresholds. Legal, court room specific tests such as the Daubert test for the admissibility of evidence, which was developed to influence legal digital forensics frameworks, methods, and tools have also, though archival adoption, influenced archival digital forensics. The Daubert test dates back to as far as 1923 (case Fry and Daubert, 1923 Court of Appeal in the District of Columbia, Daubert vs. Merill Dow Pharmaceuticals, 1993), in which the court ruled to defer to "general acceptance" of standards within the forensics community, and then "testability," over their previous reliance on peer-reviewed acceptance of forensic methods. The Daubert test then

---

[22] Charles T. Cullen. "Authentication of Digital Objects," in "Authenticity in a Digital Environment," Washington D.C., Council on Library and Information Resources, May 2000, 5 http://www.clir.org/pubs/reports/pub92/cullen.html, "Leaving us "at the mercy of catalogers.

[23] As quoted in Cullen, "Authentication of Digital Objects," 11.

[24] Data collected and verified during field work.

[25] Jeremy John Leighton, "Digital Forensics and Preservation," 13

allowed for the notion of admissible forensic evidence based on these criteria rather than placing the burden on the court. These legal thresholds have also, of course, impacted archival workflows to, in turn, be overly conservative and granular.[26] As a result of this dependence, archival workflows are hindered.[27]

Predating both fields by centuries, the seventeenth-century field of scholarly "diplomatics" has, more recently been proposed for adaptation to digital archives frameworks. Lucian Duranti advocates that "diplomatists were (in fact) the forensic scientists of their day;" they authenticated records "in a court of law when the rights they attested to were challenged and their trustworthiness as records questioned."[28] Diplomatics was formerly relegated to the authentication of historical documents, but its development underpins the development of historiography, legal theory, and the concept of ″evidence″ itself. Additionally, its focus on procedures, or on ″conventions, protocols and formulae that have been used by document creators," increases understanding of born digital document transmission, and distinguishes between the facts which the documents purport to record and the actual truth of the facts themselves.[29] Duranti argues that these aforementioned strengths suit it uniquely to address born digital record keeping. Duranti puts forth the need to integrate concepts from ″modern″ diplomatics and forensics to create a new discipline better suited to ensure trustworthiness, "Digital Records Forensics." Her series "New Uses for an Old Science" make the case of the complementary aspects between the two fields

---

[26] Brian Carrier, "Open Source Digital Forensics Tools: The Legal Argument," October 2002, accessed August 15, 2014, http://www.digital-evidence.org/papers//opensrc_legal.pdf.
Also in Johns, "Digital Forensics and Preservation," 24.

[27] Christopher V. Marsico, "Computer Evidence v. Daubert: The Coming Conflict," (student paper, Purdue University School of Technology, 2004), https://www.cerias.purdue.edu/assets/pdf/bibtex_archive/2005-17.pdf.

[28] Luciano Duranti, "New Uses for an Old Science," *Archivaria*, 28 (1995), http://journals.sfu.ca/archivar/index.php/archivaria/article/viewFile/11567/12513, 43.

[29] Duranti, "New Uses," 17-18.

today and argue that in fact, through adapting this ancient field, the archival sciences will be better prepared to discuss and address digital preservation challenges.

Workflow

This next section goes into some depth with workflow concerns and steps unique to adapting ″digital forensics″ in archives. In addition to the frameworks and tool sets, then, archives implementing digital forensics must adapt digital forensics practices into their general digital preservation workflow. Legal digital forensics workflows do not have the same provisions for long-term preservation and access. Nor do they not focus on legacy media. The complexities and variability in workflow, however, are high barriers to entry for archives. Differences in media type, copyright and intellectual restrictions, and interpreted content all may necessitate case-specific workflows. Workflows within the same institutions change significantly depending on such collection characteristics. The diagram below can illustrate a simplified version of a generic digital forensic workflow in an archive.

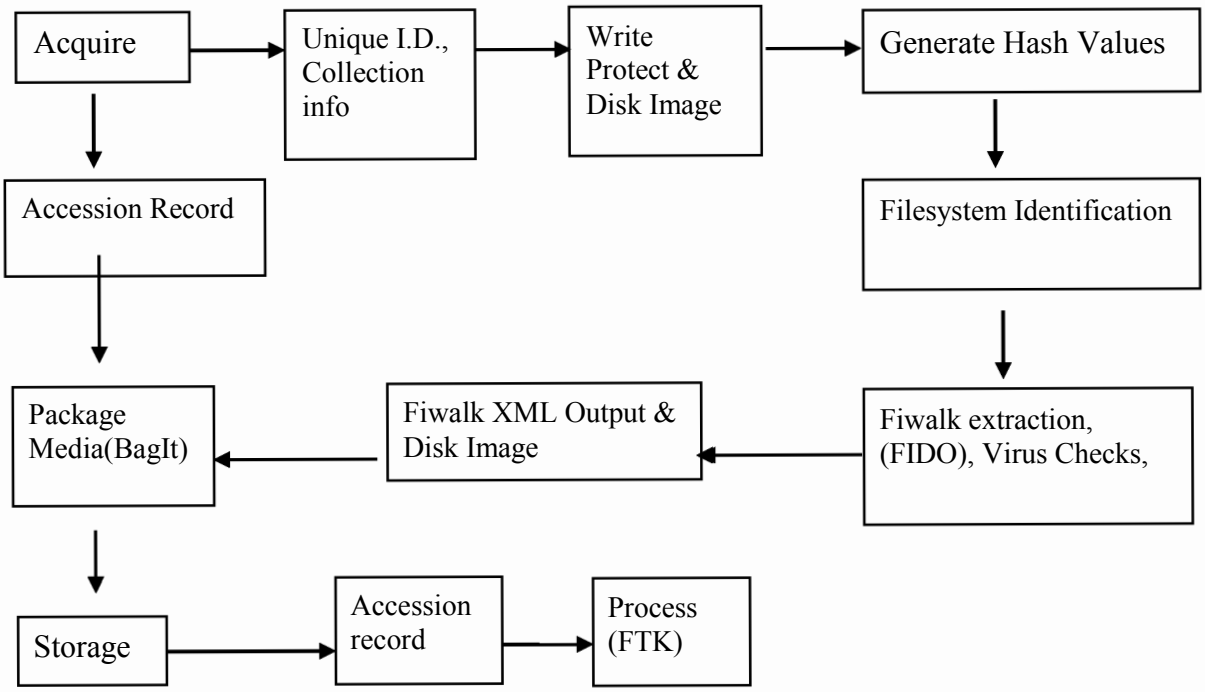


Figure 3: Digital Forensics Workflow (Mark Matienzo "Fi-Walk with Me")[30]

Archives need to understand how digital forensics will be integrated in an overall digital preservation workflow; some of the workflow steps above are already common to digital preservation. Integration presents many challenges, however. I will address these in the next section.

Workflow Adaptation Concerns:

•Pre-accessioning

More than ever, technical, legal, and privacy complexities compel archives to prepare before acquisition. This necessity involves much more than maintaining a chain-of-custody. Archives must survey collections prior to ingest to assess and plan for technological dependencies necessary for media preservation, such as file formats, hardware, and software.[31] Many archives may discover some part of the workflow needs addressing; they may not yet have the capabilities for preservation and researcher access. Or an archive may discover that the collection would require too many resources with the restrictions on the material and may need to re-negotiate accessioning. Lastly, privacy

[30] Mark Matienzo, "Fiwalk with Me: Building Emergent Pre-Ingest Workflows for Digital Archival Records using Open Source Forensics Software," (presentation at Code4Lib 2011, Bloomington, IN, February 7-10, 2011) http://matienzo.org/storage/2011/2011Feb-Code4lib.pdf.

[31] Gabriela Redwine, Megan Barnard, Kate Donovan, Erika Farr, Michael Forstrom, Will Hansen, Jeremy Leighton John, Nancy Kuhl, Seth Shaw, and Susan Thomas, "Born-Digital: Guidance for Donors, Dealers, and Archival Repositories," Council on Library and Information Resources, October 2013, clir.org/pubs/reports/pub159.

norms, especially with online access, have yet to be established.[32]

•Disk Image

Often conflated with digital forensics, disk imaging is a key contribution to workflows. It frees one from any dependence on the failing removable media by not only making a preservation copy, but a copy so exact that calls into question the very notion of originality. It also accomplishes many other purposes simultaneously: it refreshes the media, creates a bit-exact clone of the bit stream, creates a more actionable format, and creates a "sandbox" for further curatorial experimentation. This step goes "below" logical structures in its capture and yet, it is relatively simple. This ability to access and exactly clone material is a major contribution to archival workflows.

While creating an exact clone frees one from making permanent curatorial decisions, disk imaging may be unnecessary and/or unsupportable for an archive if: 1) the files were not meaningfully arranged (ex: if the files were copied for transfer); 2) the contents are well-documented; and/or 3) the archive does not have a preservation workflow to support disk image files.

It may be that the chain-of-custody and provenance have already been broken; files may be haphazardly copied onto media for the purpose of transferring to an archive. The resulting file arrangement and accompanying metadata has already altered irrevocably. The time-date stamps marking file access and editing times, for example, will have changed. Relationships between files and directories are also no longer preserved.

The seemingly indisputable argument in favor creating disk images is that since

---

[32] Kam Woods and Christopher A. Lee, "Acquisition and Processing of Disk Images to Further Archival Goals," (Proceedings of Archiving 2012. Springfield, VA, Society for Imaging Science and Technology, 147-152.), digpres.com/publications/archiving-2012-lee-woods.pdf.

there are limited reads to media before failure, why not? The alternative is an irrevocable decision. The extra time and effort for less dense media is minimal and the effort is relatively simple.

One concern that prevents the wholesale adoption of disk imaging is the threats to donor privacy. Archives can choose to image forensically and include deleted content or image logically to exclude such content. A complete disk image not only allows one to not only recover hidden files and incompletely over-written files or deleted files as well as password protected files. Today's standard donor agreements do not take into account this added possibility (yet). Disk images also involve decision-making on the part of archives as to what type of files to output and involve some sacrifice; some file types are proprietary, are only supported by some software common to forensic toolsets, or are only comparable to some media types.

•File system Identification[33]

File systems structure, organize, and store data on media for recall. An operating system will support one or more file system: MS Windows supports NTFS, FAT32; Apple Mac Operating Systems support HFS, HFS+; and Linux Operating Systems support ext-4. Optical Media commonly uses ISO 9660 and UDF. Legacy file systems include Amiga, OFS, FFS, OS/2 HPFS.[34] This step, file system identification, carries an assumption that the individual has a deep understanding of legacy media and computing to deduce interoperable issues with file systems. Many tools along the digital forensics workflow support some files systems, but not all of them or not all of them completely.[35]

---

[33] File system, Volume, Partition, and Operating System are grouped together here.

[34] Gareth Knight, p. 43.

[35] Don Mennerich and Mark Matienzo, "Pitfalls! Working with Legacy Born-Digital Media in Special Collections," (presentation at Code4Lib 2013, Chicago, IL, 2013) http://matienzo.org/storage/2013/2013Feb-code4lib-pitfall.pdf.

•Extractions, de-duplication, redaction

Many born-digital files may be duplications or non-authored irrelevant content (such as MS Office), depending on the curatorial conception of the collection. Additionally, privacy concerns inherent in any media type, analog or digital, proliferate with the current massiveness of digital collections. Addresses, social security numbers, credit card information, and the like should be restricted, but the shift to email and personal computing complicate privacy standards today. What constitutes restricted content is very much specific to each collection. This step of the workflow is a major impediment to processing and access. Comparing signatures with national registries and running similarity checks with tools such as SDHash help to streamline this time-consuming process.

•Arrangement and Description

Archivists will need to draw boundaries delineating what constitutes each collection, content, and items. These broader, networked, and complex materials question traditional methods of intellectual arrangement and finding aid organization.[36] There are many ways in which archives intellectual interpretation of complex born digital collections challenges traditionally procedures and foundational archival procedures.[37] The baseline unit is not standardized: is it a file? A document? A disk? Transposing meaningful complex hierarchical relationships of component born digital items may also be challenged. The complexity offered in many born digital materials, however, seems a great mismatch to traditional arrangement and description

•Researcher Access

---

[36] Erika Farr, in conversation with author, June 11, 2014.

[37] Jefferson Bailey, "Disrespect des Fonds: Rethinking Arrangement and Description in Born-Digital Archives," *Archive Journal* 3 (Summer 2013), http://www.archivejournal.net/issue/3/archives-remixed/disrespect-des-fonds-rethinking-arrangement-and-description-in-born-digital-archives/.

Few complex media collections are accessible. Researcher access itself presents wholly new hurdles in the workflow.[38] Making a collection accessible could necessitate an entirely separate workflow based on collection characteristics. For example, concerns such as ROM (read only memory) chip rights, file format registries, and the essential behaviors and characteristics of a work dictate access levels and types, whether they are migrated, emulated, online-accessible, or restricted to in the reading room, for example. Hybrid complex collections, such as the ones included in this study, may need to tier access to include: a virtual reading room accessible online, emulated or migrated environment accessible on-site, a text-searchable database, and a finding aid.

Workflow integration, then, necessitates significant work by archives. While there are many different tools, the questions are not just technological. They involve questions that invoke ethical quandaries and responsibilities to both the donors as well as future researchers. They involve interrogating major archival principles and, most mundanely, involve significant administration and adaptation depending on collection characteristics. Both AIMS, or "An Inter-institutional Model for Stewardship" white paper (2009-2011) and Martin Gengenbach's 2013 thesis, however, provide some in-depth workflow documentation from research universities and cultural institutions.[39] By documenting eight workflows, Gengenbach illustrates some of the different possible paths necessary with different legacy media. Issues such as media type (optical, hard disk, or floppy, for example), dictated different initial paths before re-converging in.[40]

---

[39] "AIMS Born-Digital Collections: An Inter-Institutional Model for Stewardship," January 2012, http://www.digitalcurationservices.org/files/2013/02/AIMS_final.pdf.

[40] Gengenbach. "The Way We Do It." His work has through documentation of workflow processes, decisions, and documents.

Toolsets

Digital forensics is strongly associated with obsolete media: photographs of floppy disks and 1980s era personal computers. The *materials* studied as well as used-that is the hardware and software are novelties requiring extensive explanation. Forensics analysis must involve the introduction of a host of complex toolsets for different points in a workflow for different media types. Forensics toolsets are "a set of software tools." A forensic workstation can have a variety of different tools available that will include imaging hardware/software, write-blocking hardware/software, and even, in some cases, older computer environments to open and evaluate the media. Until recently, adapted tools required familiarity with command-line and did not include graphical user interfaces. Ricky Erway and the Online Computer Library Center (OCLC) have contributed several publications introducing these tools (and steps) to the archives and library world; her work serves as a concise introductory overview of common workflow tools.[41] While archivists may look to find solutions with a single list of open-sourced tools, as of now there is no simple solution. Before designing a workstation and purchasing tools, one must account not only for the department′s own technical support and proficiency, but collection material format types and related dependencies such as physical connectors, processing speed of connections, operating system support, and original purpose and use, to name a few.

These tremendous complexities influence many institutions to purchase a Forensic Recovery of Evidence Device (F.R.E.D.), a suite of high-powered computers outfitted with many

---

[41] 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) http://digitalpowrr.niu.edu/wp-content/uploads/2012/09/Resources-for-Technical-Steps.pdf. Also see: Julianna Barrera-Gomez and Ricky Erway, "Walk This Way: Detailed Steps for Transferring Born-Digital Content from Media You Can Read In-House," OCLC Research (2012), oclc.org/content/dam/research/publications/library/2013/2013-02.pdf.

connecting cables, bays, and drives; preloaded with forensic software. While built initially for law enforcement divisions, F.R.E.D. simplifies technical complications with its many built-in features and the availability of technical support.

The removable media under review may itself also present significant restrictions. Different hardware itself may require different obsolete connectors, drives, and hardware in order to read the media. These depend on the type of carrier media: a 3½ inch floppy disk will require very different hardware and drivers for a modern computer to be able to read and image its contents. Additionally, as the following at-risk media types: 3½ inch floppy and optical media examples illustrate, working with each of these legacy media hardware types necessitate an understanding of the diversity of its possible uses in order to most successfully read and preserve the content:

> **3½ inch floppy disk** (late 1980s – 2000s)[42]
>
> Factors for preservation include data density (double-density, high density, 80 KB, 1.44MB), single-sided/double-sided, file systems (FAT, MFM, HFS), operating systems, disk geometry (tracks, sectors, byte offsets), and disk rotational speed. Floppy disks have built-in write-protection tabs.[43] ″Bad″ reads necessitate adjusting variables above, trying different computers, and re-reading the disk several times. Another beneficent attribute is that in the case that the file system identification is unsuccessful, tools (e.g. the KryoFlux disk controllers) can still image the magnetic flux transition for future curatorial intervention, operating well

[42] Guerrero, "Removable Media," 5-10

[43] "List of Floppy Disk Formats," Wikipedia, http://en.wikipedia.org/wiki/List_of_floppy_disk_format

below the logical organization of digital bits and treating the disk like analog magnetic media.[44]

Optical Media (ex: DVD, CD-ROM, CD):

Optical media can involve significant structural variance affecting preservation in both physical components, such as laminates, dyes, and layers; and logical structures. Until recently, optical media was recommended as a removable media in a preservation workflows; it is now deprecated. The variations too often lead to total read failures. Optical Media does not require a write-blocker usually; writing onto optical media requires specific write commands. Reading media with incompatible systems may readily result in successfully mounting and deceptive partial content viewing. Common Optical Media file system types include: ISO 9660 (with extensions like Joliet, Rock Ridge), Universal Disk Format (UDF), and Compact Disk Digital Audio (CD-DA). ISO 9660 is used with higher level add-ons, such as Joliet and HFS that provide greater functionality.[45] UDF (created to replace ISO9660) has been adopted by the DVD Consortium and is commonly used for DVD-ROM, DVD-video, consumer CD-R/RW.[46] CD-DA is used with audio and does not contain discrete files; it instead captures audio waveforms with linear-pulse code modulation. The differences between these three formats present

---

[44] "Kryoflux," The Software Preservation Society, http://www.kryoflux.com/.

[45] Alexander Duryee, "An Introduction to Optical Media Preservation," *Code4Lib Journal* 24 (April 2014), http://journal.code4lib.org/articles/9581.

[46] "Universal Disk Format," Wikipedia, http://en.wikipedia.org/wiki/Universal_Disk_Format. Also see, "CD-ROM Drive May not be able to read a UDF-formatted disc in Windows XP," http://support.microsoft.com/kb/321640/en-us. "Wenguang's Introduction to Universal Disk Format," https://sites.google.com/site/udfintro/.

different preservation workflows. Imaging the media may forestall mis-identifications and loss of "hidden" material only seen with a by looking through directory structures.

The nomenclature of digital forensics has become well known, in part because of the efforts to encourage adaption of forensics on the part of the BitCurator project, a multi-institutional two-year project aimed at packaging open-source micro-tools into a suite complete with graphical user interface. The suite includes a long list of micro-processing tools that can be swapped in and out depending on need and include many tools such as: ClamAV (a virus scanning software), Nautilus Scripts, and "The BitCurator Reporting Tool" (a tool that sequences various forensics tools and outputs reports).[47] Toolsets must, however, go beyond accessing media types. The below table includes just some of the tools in use today.

Tools	Use	Properties	Adoption
Guymager[48]	Disk imaging utility. Also includes imaging process metadata.	Free, open source	Part of BitCurator suite of tools.
Bulk extractor[49]	Security scanner. Looks for social security, email addresses, etc…	Free, open source	Part of BitCurator suite of tools.
The Forensic Toolkit (FTK)[50]	Commercial software allowing robust analysis inc. batch similarity identification of files and file previewing. Can process HFS, HFS+, AFF, can parse NTFS, FAT, EXT, ReiserFS, VxFS, UFS1, UFS2, CDFS, ….	Greater file format support, Does not generate DFXML, $4000	Columbia University[51], New York University, Stanford University[52]

[47] BitCurator, http://wiki.BitCurator.net/index.php?title=Software#Tools_in_the_BitCurator_Environment.
[48] "Guymager homepage," http://guymager.sourceforge.net/.
[49] "Bulk Extractor," Forensics Wiki, forensicswiki.org/wiki/Bulk_Extractor.
[50] "Forensic Toolkit," Forensics Wiki, http://en.wikipedia.org/wiki/Forensic_Toolkit.
[51] Jane Gorjevsky and Dina Sokolova, "Infrastructure Development: Multiple Digital Content Types in a Single Collection," (presentation at Digital Preservation 2014, Washington, DC, July 2014).
[52] "Stanford University Born-Digital Forensics Lab," Stanford University Libraries, http://library.stanford.edu/research/digitization-services/labs/born-digital-forensics-lab.

FTK Imager	Disk imaging tool. Retains metadata about imaging process. Reads: Raw, SMART, ISO, AFF, Encase	Free	Proprietary
Forensic Recovery of Evidence Device (F.R.E.D.)[53]	Suite of computer workstations with accessible ports and drives for safe access	$5000	Columbia University, New York Public Library, Emory University, Stanford University, University of North Carolina,
KryoFlux Disk System	USB floppy disk controller, images flux transitions, format agnostic. Supports MFM and Apple macintosh sector encoding	Cost varies for individuals (inexpensive, approx. $125) to institutions (expensive, or approx. $4000)[54]	No current alternative[55]
The Sleuth Kit (TSK)[56]	Multi-tool investigating file system analysis (supports NTFS, ISO 9660, HFS+, FAT12, FAT16, FAT32, ext2, ext3, UFS1, UFS2)	Free, open source	Part of BitCurator suite of tools.
FIDO, DROID, TIKA	Different file identification registry software. [57] Many use PRONOM data.	Free, open source	These are different types of tools for the overall purpose of file identification.
Tableau Write-Blocker[58]	Hardware, "forensic bridge" that prevents writing of media, time-stamp alteration	$300-Varies	While there are software alternatives, hardware is the prudent choice.

Figure 2: Common Tools[59]

As of now, the tools are "not there" yet in the archival world. As for all digital forensics components, the tools are still undergoing adaptation for more successful implementation in an

53 "F.R.E.D." Digital Intelligence, digitalintelligence.com/products/fred/.

54 Prices are not disclosed online. These figures are from discussion with Donald Mennerich.

55 CatWeasel, another disk controller, has been discontinued.

56 "The Sleuth Kit," sleuthkit.org/.

57 Gary McGath, "The Format Registry Problem," *Code4Lib Journal* 19(January 2013), http://journal.code4lib.org/articles/8029.

58 "Forensic Bridges," Tableau, guidancesoftware.com/products/Pages/tableau/products/forensic-bridges.aspx.

59 Other resources for this list include, "the Forensics Wiki Tools," http://forensicswiki.org/wiki/Tools, and "How do these tools address archival concerns," BitCurator, http://wiki.bitcurator.net/index.php?title=How_do_these_tools_address_archival_concerns%3F, and the Appendix in Gengenbach, "The Way We Do It."

archival setting. Many available digital forensics tools are made specifically for criminal investigation, and so they do not necessarily work well with legacy material, for example.

Investigative digital forensics is focused on contemporary computing environments; there is no long term preservation or access mandate. For archives to successfully adapt digital forensics methods, then, there is a need to find ways to work with legacy materials, to think about access, and to address intellectual property and privacy concerns.

This is a cursory listing of the many issues raised with digital forensics. This does not signify, however, that these barriers to digital forensics adaption are too much for smaller institutions.[60] While "lack of funding, time, and expertise" may exclude many from digital forensics, recent complex media collections processed at Getty Institutional Records and the Manuscripts Division of the Stanford University Libraries Department of Special Collections and University Archives illustrate the development of workflows and strategies when there was extremely limited funding, time, and expertise.[61]

---

[60] Laura Wilsey, Rebecca Skirvin, "Capturing and Processing Born-Digital Files in the STOP AIDS Project Records: A Case Study." *Journal of Western Archives* 4:1(2013), http://digitalcommons.usu.edu/westernarchives/vol4/iss1/1.

[61] Cyndi Shein, "From Accession to Access: A Born-Digital Materials Case Study," *Journal of Western Archives* 5:1(2014), http://digitalcommons.usu.edu/westernarchives/vol5/iss1/1. Also, Gorgevsky, "Infrastructure Development."

1.2 Materiality and Format

Every contact leaves a trace -Edward Locard (1877-1966) [62]

Data flows everywhere around us. Such ubiquity renders it invisible. The daily engagement with digital materiality on our phones, tablets, and computers paradoxically conjures the illusion of data's immateriality. The seeming intangibility is heightened with the growing reliance on cloud storage and our newfound ability to access data anytime, anywhere. Yet the feasibility of digital forensics is founded very much on the opposite notion: data is material. "There is no computation," computing forefather John von Neumann notes, "without data's representation in corresponding physical substratum."[63] Date must have a physical correlate in computing. This fact is can be easily overlooked by archivists first comprehending digital forensics activities. Archivists must then grapple as much with this physical correlate as with the notions embedded in the data. If this were not the case, archivists would have an impossible task: preserving pure ephemera[64].

A similarly mistaken preconception/simplification is that data is highly volatile. The volatility of data is, in fact, the organizing principle in forensic investigations in which areas higher on the "order of volatility" are prioritized because of their high risk for corruption, deletion, or destruction for investigation.[65] Data is volatile, but within the forensics field, its remanance is just as often stressed. So-called "erased" data may still be available until they are thorough written over with new data. Both are accounted for characteristics stressed in forensics investigation. Both characteristics are of preeminent concern; these paradoxical characteristics are preconditions

---

[62] Leighton, "Digital Forensics and Preservation," 13.

[63] As quoted in, Kirschenbaum, Matthew G. "Mechanisms: New Media and the Forensic Imagination. MIT Press (2008), 27. Much of this section is indebted to Kirschenbaum's agenda-setting work.

[64] Kirschenbaum, "Mechanisms."

[65] Examples of OOV: Microprocessor registers, microprocessor cache, RAM, Hard Drive, peripheral memory (R/W), Write once. uio.no/studier/emner/matnat/ifi/INF3510/v12/learningdocs/WS-INF3510-2012-L11-QA.pdf

to modern computing. Conversely, ensuring that data is securely destroyed is also difficult. From the perspective of government security, the list of activities for secure data removal is so extreme as to sound ludicrous: run over items with tanks, rip them apart, de-gause them with a military magnet, incinerate them.[66] Once data is inscribed, deletion is not a keystroke. Rather, secure disposal involves multiple processes. If data is online, then it can be even more difficult to "erase." As Michael Calonyiddes says, "electronic mail and computer records are far more permanent than any piece of paper."[67] The issues behind both data recovery and deletion, then, are more a matter of costs. Technicians can work with material at the micron level. Their ability to engage with media at that level allows them to possibly recover it. While the expertise and skill required will be out of scope for most archives, it is important to note the complexities of data.

Digital Archaeology, Media Archaeology, and Rosetta Computers

While digital forensics's roots in computer science and law enforcement continue to have major influence on the field, less known and acknowledged is the influence from other fields. Not only does digital forensics provide new frameworks, tools, and workflows, its various methods raise interesting hermeneutical questions for this new technological era.[68] As such, digital forensics is as indebted to literary theory and media studies for new approaches to understanding legacy material. Of special interest are the emerging fields of digital and media archaeology and the approaches they offer to analyzing and interrogating complex work. The challenges with complex media, conversely, necessitate cross-disciplinary collaboration with other disciplines.

---

[66] Kirschenbaum, "Mechanisms," 26.

[67] Michael Calonyiddes, Computer Forensics and Privacy (Norwood, MA: Artech House, 2001), 4. As quoted in Kirschenbaum, Mechanisms, 51.

[68] Kirschenbaum, "Digital Forensics and Born-Digital," 2.

Archivists' use of digital forensics, then, is a broad umbrella term with various connotations; similarly "digital archaeology" is also much used across disciplines without clarity. For the purpose of this work, we will consistently use the term digital forensics because of its wide-spread adoption, but digital archaeology merits further exploration. Archaeology is defined in the Oxford English Dictionary as "the study of human history and prehistory through the excavation of sites and the analysis of artifacts and other physical remains."[69] Digital archeology, is defined by Dan Farmer and Wietse Venema as "about the direct effects from user activity, such as file contents, file access time stamps, information from deleted files, and network flow logs," with the suggestion of further adoption of this term. Farmer and Venema, pioneering figures in the computer security digital forensics world, use both digital forensics and digital archaeology in the influential *Forensic Discovery* (2005). They acknowledge that the literature can use the terms interchangeably, but point to some ways in which archaeology is the more evocative and true to the process of forensics analysis. In their descriptions of data remanance, they provide an analogy: deleted file information is like a fossil in that its skeleton may be missing a bone here or there, but the fossil remains, unchanged, until it is completely overwritten. As one uncovers layer upon layer, "information becomes more and more accurate because it has undergone less and less processing. But as we descend closer and closer to the level of raw bits the information (conversely) becomes less meaningful, because we know less and less about its purpose."[70] Thus even as we go closer, bypassing organizational structures, the data *seems* more abstract. Archaeology is also used in early work by Ross and Gow (1999) to describe data recover. Leighton suggests limiting the term "for the situation where fragments of information

---

[69] Oxford English Dictionary. 2nd ed. 20 vols. Oxford: Oxford University Press, 1989. oed.com/.

[70] Dan Farmer and Wietse Venema, *Forensic Discovery*, (New York: Knopf, 2007), 9, 59-51, (as cited in Kirschenbaum, Digital Forensics and Born-Digital Preservation, p7)

are not only recovered but used to investigate and interpret social circumstances, such as online communities or early computer game players."[71] This calls into mind the necessary interpretation of remnants to recreate some sort of narrative or timeline of activity. For the purpose of this work then, we will restrict our usage to sociological "interpretations as we dig through layers."[72]

"Cultural studies, film, media arts, and history – it seems no disciplinary bounds can hold media archaeology."[73] Archival uses of the term attach some of these developments. Media archaeology can be defined very broadly as "an approach to the social history of technical media" that is decidedly non-linear and non-historicist.[74] The use of archaeology has roots in the Foucaultian notion of "the archaeology of knowledge," and has been used frequently in film studies to explain major technological change (i.e. the coming of sound and proto-cinema). Laurent Mannoni's championing of "the archaeology of cinema" as the study of proto-filmic technologies is another example, linked to cinema studies adoption and adaptation of "archaeology." A commonality is the emphasis on the mediation of materiality in bits, code, and the experiential slow flickering of screens. However vague and all-purposeful, media archaeology has also come to represent the "rubbish bins of history and the resuscitation of the obscure.[75] In English literature, the field of media archeology has been championed by Lori

---

[71] Leighton, "Digital Forensics and Preservation," 22. A conversation of competing digital archaeology definitions in the Digital Preservation community is available: Leslie Johnston, "Digital Archaeology," *The Signal: Digital Preservation*, September 10, 2013, http://blogs.loc.gov/digitalpreservation/2013/09/digital-arhcaeology/.

[72] Charles Henry, Preface, Preface to "Digital Forensics," by Matthew Kirschenbaum, et al. (2010) or as he put it, "tunneling through layer upon layer of abstraction."

[73] Dr Scott Anthony, "What is Media Archaeology?" review of "What is Media Archaeology?" by Jussi Parikka, *Reviews in History*, November 2012, http://www.history.ac.uk/reviews/review/1343.

[74] John Armitage, "From Discourse Networks to Cultural Mathematics: An Interview with Freiderich Kittler," *Theory, Culture, Society* 23: 17 (2006):32, DOI: 10.1177/0263276406069880. Kittler also finds the term problematic to define. Jussi Parikka's synopsis and synthesis of media archaeology is also helpful. "What is Media Archaeology," Cartographies of Media Archaeology blog, http://mediacartographies.blogspot.com/2010/10/what-is-media-archaeology-beta.html.

[75] Anthony, "What is Media Archaeology?"

Emerson and Lisa Gitelman (among others). Lori Emerson's approach has been particularly fruitful for archivists; it acknowledges the materiality of obsolete media with her collections of "failed" and obsolete computers. Emerson goes further; she has collaborated with other institutions, granting access to the Media Archaeology Lab, a curated collection of obsolete computers, peripheral hardware, operating systems and software housed at the University of Colorado at Boulder's Media Lab. Created in 2009, the Media Archaeology Lab gives researchers a rare opportunity to experience obsolete technologies, or "failures of history," as she has sympathetically termed them, first-hand.[76] As libraries and archives work to preserve obsolete born digital media, they have been able to adapt some of these approaches for their own purposes. One of the earliest uses of "archaeology" with digital forensics began with the pioneering establishment of the University of Texas-Austin's "Digital Archaeology Lab,"[77] and its concomitant processing of Michael Joyce hypertext literature at the Harry Ransom Center (received in 2005).

Mediating hardware, so-called "Rosetta Computers," have also become more common as translation mechanisms to access legacy media. Rosetta Computers are so-named because of their function as bridging different periods of technological history. As the term may indicate, a Rosetta Computer will, for example, have accessible ports that serve as a bridge or conduit connecting older and newer machines together that do not themselves have the requisite access points.[78] They may have DB-25 parallel ports, RS-232 serial ports, or, for example, consist of a machine with a floppy disk drive.[79] A concrete example of this is the Mac PowerBook G3.

---

[76] Lori Emerson, "Media Archaeology Lab," University of Colorado Boulder, http://loriemerson.net/media-archaeology-lab/.
[77] "Digital Archaeology Lab," University of Texas Austin, https://www.ischool.utexas.edu/about/labs/digital_archaeology_lab.
[78] Doug Reside, "Rosetta Computers," in "Digital Forensics and Born-Digital," by Kirschenbaum et al, 19-20.
[79] Guerrero, "Removable Media," 27-28.

"The PowerBook in our lab has a SCSI port, to which we could connect the SyQuest drive, as well as a drive for Zip disks, with which we could transfer the data to a more recent machine."[80] After the G3's released in 1998, SCSI ports were no longer standard. The late 1990s, then, represent a rift in technological hardware that have major effect on archival access and preservation.

While most institutions processing obsolete material may wish to avoid collecting and maintaining obsolete machinery, older machines provide essential access to otherwise inaccessible legacy materials. In allowing access to the material, they enable archivists to arrange material, to plan for an emulation or migration, and to map essential qualities. Accessing and experiencing an object's meaning will mean, in the best of all situations, having a hand in a variety of obsolete hardware, software, and operating systems in order to experience the born digital artifact in something approaching its native environment. This approach of collecting so-called Rosetta Computers to translate and bridge major technological shifts can be a resource-effective strategy. While not necessarily a standard, recommended practice, looking to the past for accessing and preserving things for the future is something that media archaeology contributes to archives today.

While technologists may dismiss this section's interdisciplinary foray as mere "semantics," it is important to note not only the overlap in the meaning of terms, but the pragmatism for the archival field in maintaining dialogue with these diverse areas of scholarship in the computer sciences, digital humanities, and literary fields.[81] There are many fruitful outputs from these traditionally separate disciplines for our field. The Maryland Institute for Technology in the

---

[80] Mark Matienzo. "Collaboration before Preservation," *Manuscripts and Archives blog*, September 26, 2013, Yale University, http://mssa.commons.yale.edu/2013/09/26/collaboration-before-preservation/.

81 While there is widespread debate, characteristics common to Digital Humanities projects are: collaborative, multidisciplinary work, creation of an online interface accessible to the public, and use of info graphics, network graphics, and other visualizations.

Humanities and its Humanities Intensive Learning and Teaching program is a longstanding formalized program, founded in 1999, nurturing multidisciplinary collaboration.

A recent addition by the Library Science community to find ways to address these issues is the "Jump In" initiative. The Online Computer Library Center (OCLC) researcher, Ricky Erway, much more recently spearheaded the "Jump In" initiative to coordinate among archival institutions to fill in the different needs and availability of obsolete hardware, software, and technical expertise among them. From her previous research, it was clear that many institutions had a particular type of media or equipment or expertise that would greatly benefit another institution.[82] Erway wanted to create an infrastructure for coordination among them for skill and tool exchanges. In the "Jump In" initiative, archival institutions themselves may be able to fill in the technology and competency gaps for one another in localized and regional networks for exchange. For example, an institution may need to outsource the imaging of a collection and another institution may have the capacity to chip in. While the "Jump In" initiative is in its nascence, it is clear that unless smaller collecting institutions do not create more widespread support for their work, the work will not necessarily get done later. Non-profit cultural institutions will not be able to rely on market driven third-party services. The market share for archival services with obsolete, non-contemporary media is too low to ever be cost-effective. Waiting to out-source to vendors may not be the option. Costing would render any such services unaffordable; cultural institutions will have to rely on their own resources and ingenuity.[83] Erway's scholarship and the creation of these different workshops point to attempts within the community to address some of the need within the community.

---

[82] Erway, "Walk This Way" and "You've Got to Walk."

83 Ricky Erway, conversation with author, March 24, 2014.

Through applying techniques from other disciplines and adapting them collaboratively, digital forensics offers an archaeological approach to "dig" through previously inaccessible layers of meaning.

1.3 What Is 'It'

Complex media is defined broadly to include "file types in different formats, applications to coordinate the files, and operating systems to run the applications."[84] Complex media has multiple dependencies and each of these components must be accounted for. If one component fails, the complex media may no longer open or function.[85] With all of these components, it is necessary to determine "essential properties" for prioritizing. Determinations may be motivated by technical constraints and projected research applications, for example. While much time is spent discussing technological challenges, the conceptual issues at stake in answering "what is it?" for preservation and access are equally valuable.

Prioritizing essential elements at the outset and making the intellectual distinctions regarding the object in question should drive preservation actions. A hard drive partitioned to include different directories of individual files and documents, as well as a database, games, or cached email make the challenges of this immediately clear: is this a singular intellectual item or many different items? Should each digital file be treated independently? Is the database static? Is the email's text sufficient, or must it be accessed through email software? The determinations can be endless when looking at complex interactive material; analogies to paper-based processing break down. While there is no consensus on how best to answer these questions, archives have been forced to make these decisions on a case-by-case basis. Existing case studies have adapted to specific collection characteristics. Different archives may prioritize different properties. While we cannot change the idiosyncratic variety in complex born digital archival collections, we should be able to standardize how to talk about and prioritize elements for preservation workflows.

---

[84] Oya Rieger, "Preservation and Access Frameworks for Digital Art Objects," (grant proposal, National Endowment for the Humanities, May 2012) 2-3, http://www.neh.gov/files/grants/cornell_universitypreservation_and_access_framework_for_digital_art_objects.pdf,

Should each "electronic record" be considered independently, for example? Or should the contents of an entire hard drive or disk itself get treated independently? These questions may represent a major shift in archival practices.

There are several existing frameworks for diagramming complex media. The Kenneth Thibodean tripartite structure, from the archival world, as well as the Functional Requirements for Bibliographic Records (FRBR), from the library science world, are both useful and necessary in interrogating complex media.

	Example components	**What to control**
Physical	Plastic, metal, laminates	Macro conditions
Logical	Encoding scheme	File System analysis
Conceptual	*Meaning*	*Interpretive*

Figure 4. Thibodean Tripartite Structure

The archivist Kenneth Thibodeau's oft-cited tripartite framework treats digital objects as simultaneously physical, logical, and conceptual.[86] This is a helpful starting point for interrogating layers of content.[87] Objects have multiple inheritances; they have physical form that is itself prone to degradation, decay, and loss. They also have logical organization of data, or data structures. Through these underlying components, an object coalesces to have significance or conceptual meaning for individual stakeholders, whether they are archivists or

[86] Paul Nicholls has suggested a similar framework of preservation of the aesthetic, conceptual, and physical characteristics," in Multimedia: A Management Perspective, Antone F. Albert, (ed.) (Belmont CA: Wadsworth Publishing Company, 1996) 242.

[87] See Andy Ulrich's Thibodean interpretation in, "Pressed into the Service of Cinema: Issues in Preserving the Software of Hollis Frampton and the Digital Arts Lab," The Moving Image 12:1 (2012): 18-43, http://ezproxy.library.nyu.edu:2255/ehost/pdfviewer/pdfviewer?sid=e7b214d2-3e25-4408-94b9-4e091bcca469%40sessionmgr10&vid=2&hid=10.

originating artists of the work, or researchers accessing the work. A similar framework from the cultural property world exists with aesthetic substituting logical.[88]

To understand the Thibodean framework, its uses, and its limitations with respect to complex media, this next section will go into further detail on each aspect. The *physical* aspect of acquired media, quite apart from any of the files and programs they may contain, present challenges that are well-documented and understood by the preservation community. Physically broken down into component metals, plastics and laminates, CD-ROMS may last, according to manufacturers, anywhere from five to one hundred years.[89] This statistic is, of course, in total disregard of content-specific dependencies. Many collecting institutions, however, have a good grasp of physical preservation. Institutions readily understand that in order to preserve media the most significant factors to control are macro conditions governing humidity and temperature. The physical stabilization of any inherent vice is readily achievable for many media carriers but this step is ultimately inconsequential with respect to the challenges of the logical and conceptual orders.[90] These other abstracted inheritances and their needs unfortunately may be overshadowed once items are catalogued and put into storage.

A consideration of the *logical*, the second inheritance, is an area in which digital forensics is particularly instructive for archives. With the complex media analyzed by digital forensics (such as CD-ROMS, removable hard drives, and floppy disks) data structuring on each medium and within each medium type require different preservation workflows. Digital forensics has been

---

[88] Kenneth Thibodeau, "Overview off Technological Approaches to Digital Preservation and Challenges in Coming Years," in *The State of Digital Preservation: An International Perspective*, July 2002, ISBN 1-887334-92-0. http://www.clir.org/pubs/reports/pub107/thibodeau.html.

[89] "Storage, Care and Handling of Magnetic Hard Disk Drives," Image Permanence Institute, accessed May 15, 2014, https://www.imagepermanenceinstitute.org/webfm_send/301 18943.

[90] "Classification of Standards," Image Permanence Institute, accessed August 20, 2014, https://www.imagepermanenceinstitute.org/testing/classification-of-standards. The standards for magnetic hard disks are still under development, actually.

crucial in filling in an unmet need by providing tools and workflows to better approach the difficulty of the logical and by forcing archivists and curators to begin to think much more granularly and technically about the vast swathe of legacy media. By logical then, Thibodeau intends to encompass any encoding/decoding on the hardware and any rules and logic used to store and access the data. This logical aspect is essentially what a disk image captures: an exact bit-level copy that maintains its organization. Understanding a logical object, then, may presuppose an understanding of how data is mapped, the architecture and logic of storage, writing, deletion and retrieval, and knowledge of the file systems, partitions, and files. File system analysis is a precondition to cloning, but there may be some guesswork in determining a file system of a media. Specific file system types each have unique historical uses and associations which affect workflows due to the particular restrictions they may present, such as file size limits, naming characteristics, and interoperability. Different digital forensics software programs compound this issue; each software system (the Sleuth Kit and FTK Imager, for example) support different file system types. And so a solid historical and technical understanding of file systems types associated with different production time periods and industries are necessary.

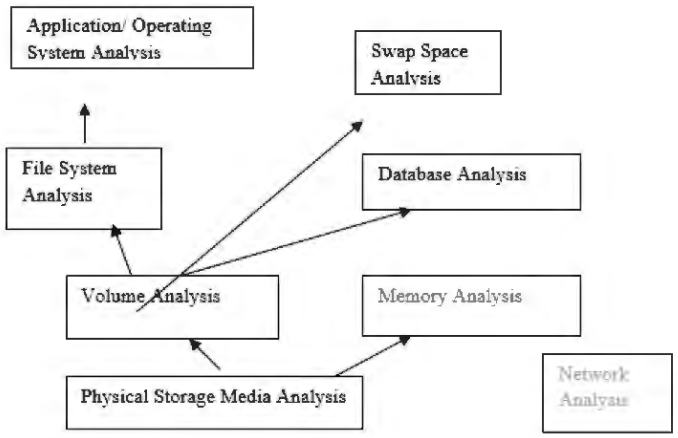


Figure 5. Layers of Organization (Brian Carrier, *File Format Analysis,* 108)[91]

Brian Carrier's diagram of forensic investigation through logical components of media is particularly helpful in conceptualizing an investigation's path through different layers of organization, from the physical media onward. As we can see in the diagram above, the logical aspect of complex media involves interrogating and navigating layers of structures inherent in a piece of media such as the volume, the file system, and the operating system. In contemporary computing, however, these logical systems are all accessed without a user's conscious awareness. The logical will also involve basic interoperable questions and determinations of the operating system used. Carrier's diagram is a simplification of the layers of embedded logical systems.

While general users of digital material may be unaware of these organizational systems, archivists working with legacy media will need to understand these components in case, as often happens, one of these layers of technology fails. In this following diagram is included to provide another method for understanding layers of logical representation inherent in born digital media.

[91] Brian Carrier, File System Forensic Analysis, (Addison-Wesley, 2005), 108. The greyed out areas are irrelevant to the digital forenisc covered in this thesis.

What, on top, is a meaningful entity is, when looked at in its unprocessed form is analog, is flux. At that state all data can be reduced to transitions, not discrete values of zeroes and ones. This is easiest to see with low-density media, such as floppy disks, with which we can simply record the flux stream if we cannot successfully gauge the file system.

Digital Resources - Levels of Representation

Level	Label	Explanation
8	Aggregation of objects	Set of objects that form an aggregation that is meaningful encountered as an entity
7	Object or package	Object composed of multiple files, each of which could also be encountered as individual files
6	In-application rendering	As rendered and encountered within a specific application
5	File through filesystem	Files encountered as discrete set of items with associate paths and file names
4	File as "raw" bitstream	Bitstream encountered as a continuous series of binary values
3	Sub-file data structure	Discrete "chunk" of data that is part of a larger file
2	Bitstream through I/O equipment	Series of 1s and 0s as accessed from the storage media using input/output hardware and software (e.g. controllers, drivers, ports, connectors)
1	Raw signal stream through I/O equipment	Stream of magnetic flux transitions or other analog electronic output read from the drive without yet interpreting the signal stream as a set of discrete values (i.e. not treated as a digital bitstream that can be directly read by the host computer)
0	Bitstream on physical medium	Physical properties of the storage medium that are interpreted as bitstreams at Level 1

Figure 6 (Kam Woods and Porter Olsen, "Digital Forensics" presentation)[92]

Last in Thibodeau's conceptual framework of the tripartite "object" is the *conceptual* object, or the meaning derived. The component obsolete hardware and software/file formats of complex media could preserve aspects of a piece, but distinguishing and prioritizing the essential

[92]Kam Woods and Porter Olsen, Born-Digital Forensics presentation (Maryland Institute of Technology in the Humanities, Humanities Intensive Learning and Teaching, "Born-Digital Forensics" class, College Park, August 2014).

meaning in a systematic way is something that the archival field is still working on. As will be seen in the case studies included, determining the conceptual aspect of complex media can include consulting primary stakeholders such as the artists-creators of the work and the anticipated researcher audience. The meaning of a work is not necessarily clear; any preservation plan will prioritize preserving different aspects of a work. While standards and technical limitations minimize the influence of whim and subjectivity, determining the conceptual should be standardized. In determining conceptual essentials, however, the unidirectionality of Thibodeau's tripartite is challenged. To recall Figure 4, we cannot actually assume that we make decisions moving from physical to logical to conceptual. The decisions regarding the conceptual may determine the physical qualities preserved. That is, movement in the figure is bi-directional and determined just as much from the conceptual to the physical. If, to cite a common example with complex media, a particular (CRT) screen type is emphasized as essential for researchers/viewers of a piece, that conceptual determination complicates the tripartite framework and will determine prioritization of aspects of the logical and physical. This last point emphasizes the need to determine the meaning of a work before beginning any preservation work, as it may have major impacts on the process.

The meaning and significance of a work has and can be also informed by analyzing "look and feel," a term borrowed from conservation. "Look and feel," can include diverse components comprising various aspects of a user's experience of an object: the pixel size on a screen, colors, the flicker and refresh rate, screen size, the speed of the program, the responsiveness and even the acoustics of the buttons and keys.[93] These many components to a complex object can be named,

[93] Jeff Rothenberg, "Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation," CLIR, January 1999, http://cours.ebsi.umontreal.ca/sci6116/Ressources_files/%5BCLIR%5Dpub77.pdf.

tracked, and even quantified to prevent loss of essential properties. In complex artwork conservation where "look and feel" is often evoked, these issues can be clarified with knowing what parts of a piece will be on viewable and display and what will be hidden and having a direct experience of the work. In the conservation world, CRT monitors, for example, are often prioritized as key to a piece and LCD monitor displays substitutions.[94] This same prioritization is less standardized in the archival world with monitor-dependent media.

Recent research has highlighted the paths and pitfalls of various emulation and migration strategies aiming to preserve an experience and the value of a game. This is the most fraught area of interpretation for conservators and archivists. What is essential to preserve in a complex piece of media with many dependencies? Hedstrom, Rothenberg, and many others have explored these questions, especially as they relate to video game and art preservation.[95]

While the Thibodeau tripartite conceptual framework can help us begin to understand important considerations for removable physical media and their content, there are other frameworks on which archivists and conservators have drawn in their attempts to standardize this highly interpretable area. Functional Requirements for Bibliographic Records (FRBR) is one of the alternative frameworks; it teases apart complex works into component work, expression, manifestation, and the item 'itself.' "FRBR is ontology that shows how a Work can be expressed in a variety of ways that are Manifested in particular media and mediums which are managed by individuals Items by the library."[96] What is the difference between an edition, a translation, and

---

[94] Gaby Wijers, "Ethics and Practices of Media Art Conservation, A Work-In-Progress," Packed: Centre of Expertise in Digital Heritage, packed.be/en/resources/detail/ethics_and_practices_of_media_art_conservation_a_work-in-progress_version_0/P80/.

[95] Margaret Hedstrom, Christopher A. Lee, Judith S. Olson, and Clifford Lampe, "The Old Version Flickers More": Digital Preservation from the User's Perspective," The American Archivist 69 (2006) 159-187.

[96] Jacob Nadal. "Print and Digital Preservation," January 20, 2014, http://www.jacobnadal.com/wp-content/uploads/2013/01/Nadal_Chapter29_PrintDigitalPreservation.pdf.

a manifestation? The complexity of mapping these interrelationships increases with interactive computer works. While FRBR was initially created for collection description, it has been put forward as a useful tool for mapping aspects of a work for the purpose of preservation. And while originally developed for end users, it can also be useful in creating preservation hierarchies and evaluations.[97]

FRBR: Group 1[98]

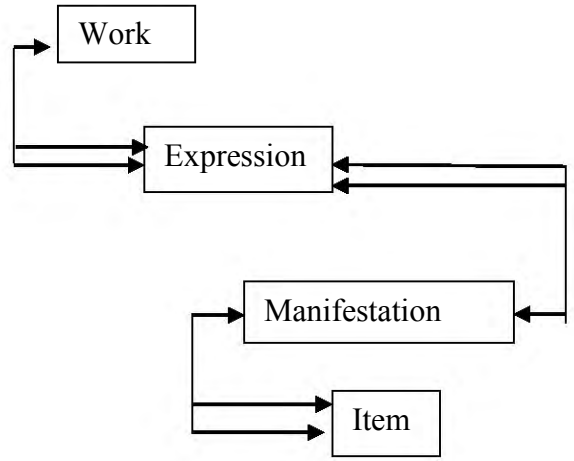


Figure 7. from "What is FRBR?" (Library of Congress)[99]=

> Work: A distinct intellectual or artistic creation.
> Expression: The intellectual or artistic realization of a work in the form of alpha-numeric, musical, or choreographic notation, sound, image, object, movement, etc., or any combination of such forms.
> Manifestation: The physical embodiment of an expression of a work.
> Item: A single exemplar of a manifestation.

In the example of the interactive CD-ROM cited earlier, then, manifestations may include all the physical CD-ROMS of the artistic work. Expressions may vary with different documented interactions. What are we trying to capture when the item is a hypertext choose your own

---

[97] Richard Newman and Ian Davis, "Expression of Core FRBR Concepts in RDF," http://vocab.org/frbr/core.html.
[98] See glossary for in-depth definitions of FRBR aspects, digitalhumanities.org/dhq/vol/4/2/000089/000089.html.
[99] Barbara Tillet, "What is FRBR? A Conceptual Model for the Bibliographic Universe," (Library of Congress Cataloguing Distribution Services brochure, dated February 2004) loc.gov/cds/downloads/FRBR.PDF.

adventure on CD-ROM? These distinctions blend together to some degree and we must think about capturing the playing of the game, the manifestation, or physical object, of the game, and whatever are the "essential" parts of the game itself. Similar to the longstanding bias and dichotomy between computer "interface" and "content," to borrow from literary and cultural studies, form and content are irretrievably enmeshed together and determine one another.

The particular computer and screen form and influence as much of the work as the CD-ROM game that is playing on it. To elaborate, CRT monitors follow the long line of projector types in a long lineage of visual aesthetics that have as much to do with the understanding of a work or piece as "the piece itself."[100] These distinctions and categories provoke a lot of discussion among textual and media scholars and among digital humanists. "While FBR provides an attractive theoretical model, the complexity of computer games as works makes its application to such software creations problematic in practice."[101] In "Twisty Little Passages," we can see that any application of FRBR to computer games is a not only time-consuming and expensive, but so complex as to make its practical application unrealistic. There are no simple solutions.

A current National Endowment for the Humanities funded project at Cornell University's Rose Goldsen Library has focused in part on this very issue, as elaborated in the April 2014 interim report on "Preservation and Access Frameworks for Digital Art Objects." They are formulating a plan to meaningfully preserve their collection of CD-ROM interactive art and will be using digital forensics tools and techniques to that end.[102] In systematically preserving their interactive CD-

---

[100] Matthew Kirschenbaum, "So the Colors Cover the Wires": Interface, Aesthetics, and Usability," in A *Companion to Digital Humanities*, ed Susan Schreibman, Ray Siemens, John Unsworth, chapter 34, Oxford: Blackwell, 2004 http://www.digitalhumanities.org/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-5-4.

[101] Jerome McDonough, Matthew Kirschenbaum, Doug Reside, Neil Fraistat, Dennis Jerz, "Twisty Little Passages All Alike: the FRBR Model to a Classic Computer," *Digital Humanities Quarterly* 4.2 (2010) digitalhumanities.org/dhq/vol/4/2/000089/000089.html.

[102] Tillet, "What is FRBR?"

ROM media, however, one of their major steps (step three) will be to systematically organize FRBR "significant properties" for interactive works in the collection.[103]

This foregoing chapter has provided basic information about various complications to understanding complex media, including how to categorize, understand, and, therefore, approach preservation. The next chapter will ground these concerns through several case studies in personal digital collections.

[103] . In terms of metadata schemas, they are also using PREMIS. "All aspects of this work (disk images, compiled emulators, ROMs, Operating Systems, virtual disks, notes, reports, etc.) will need to be preserved and will be part of this CULAR project/collection."

Chapter 2: Case Studies

In the previous chapter, I foreground digital forensics with an overview on its history, applications, and significant projects in the past several years. I have limited its definitions to those purposes most relevant for archives and archivists; I have also elaborated some of the essential steps within a digital forensic workflow and some of the intellectual and ethical complications posed by the handling of complex media. In this next chapter, I will now consider several case studies of hybrid personal collections that have, quite uniquely, been made available to researchers.

There are few published case studies in which digital forensics has been used extensively with hybrid collections containing complex, and not simply text-based, pieces. The difficulties are immense, not only for preservation but for access. Preservation and access are not simply highly technical processes; through new imperatives for online access they are potential ethical and legal minefields. These new methods and means for access (e.g., online, word-searchable databases) puts pressure on staff as they balance access with privacy and intellectual property issues. Few major institutions are able to take such risks. In the case in which no significant materials is made available online, there are still major complications of all kinds (philosophical, technical, ethical, and intellectual). The following case studies, the Salman Rushdie Papers and the Timothy Leary Papers, demonstrate some of these complications and are among a very small handful of complex material available to researchers today.

Both are two of the earliest and most well-documented and publicized test cases involving obsolete media. Other well-known complex obsolete personal papers include the 2005 acquisition of the Michael Joyce Papers by the University of Austin, Texas, and Deena Larsen's Papers at the Maryland Institute of Technology in the Humanities. As more writers deposit their

collections, the issues surrounding their digital materiality will require a more systematic approach. These refinements will be as much cultural and administrative as they will be technical. These two collections have also acquired papers by Russell Banks, John Crowley, Robert De Niro, Thomas Kinsella, Norman Mailer, and Alice Walker.[104] It is probable that few collections, however, will have the resource support to have as much complex access as the two case studies included here.

In the case of the Salman Rushdie Collection at Emory University's Manuscript, Archive, and Rare Book Library, the work on his obsolete personal computers was the first venture of its kind. Previous to the Timothy Leary collection, New York Public Library's Manuscripts and Archives Division had worked on several other born-digital collections that are accessible to researchers, notably the Gay Men's Health Crisis Record and the September 11th Fund records.

While my methodology for both case studies includes interviews and traditional research, there are several points of divergence. I was also able to work as a researcher to the Leary collection first-hand in a non-networked computer in the reading room on several occasions (four to five days) from December-January 2014. Due to my greater first-hand experience with these materials, I diverge slightly in the Leary section, elaborating more in the content and access sections. Also, while the Leary papers were opened only recently to the public, the Rushdie papers have been available to some for several years. Due to Rushdie's celebrity and the innovative work at Emory, there is much more available (if not always published) material, especially in the form of secondary sources and scholarship, on which I have been able to draw.

---

[104] Matthew Kirschenbaum, "Digital Materiality: Preserving Access to Computers as Complete Environments," http://mkirschenbaum.files.wordpress.com/2009/10/digitalmaterialityipres2009.pdf.

Chapter 2: Salman Rushdie (1947-) and Timothy Leary (1920 − 1996)

After acquisition in 2008, Emory University's Manuscripts Archives and Rare Books Library processed, preserved, and created access for Salman Rushdie's earliest personal computer for its 2014 opening ceremony. After acquisition in 2011, the New York Public Library's Manuscripts and Archives Division were able to provide access to Timothy Leary's floppy disks to researchers in 2013. Of the many unprocessed and inaccessible collections in the backlog, these two high profile collections are, then, exceptional. The case studies illuminate the difficulties in executing the access proviso, a stumbling block that I will consider in the next chapter. The complementary qualities of the two case studies; studied together, cover the spectrum of hybrid born digital personal collections. These categories include: 1. Organizational support and roles, 2. Donor involvement, 3. Intellectual property, copyright, and privacy issues, 4. Content issues, 5. Access methods, 6. Workflows. As with most of these categories, there is not yet any standardization in how collections define and delimit. Defining content, drawing the boundaries between the artifactual and behavioral are left to each archive and to each collection even. In fact, as these cases illustrate, there is no standardization across all of these categories. All the involved stakeholders-- from the donors and estates to the archives to the researchers-- do not have clearly defined roles. Major determining factors such as technological capabilities, for example, have changed things dramatically in just the past few years. This table below summarizes some of the key characteristics across the aforementioned categories:

	Salman Rushdie – Emory	**Timothy Leary – NYPL**
Organizational support and roles	Extensive, multiple	Minimal
Donor involvement	Interested, invested	Minimal
Intellectual property, copyright, and privacy issues	Elaborate, nuanced, difficult	Minimal
Content	Whole personal computing environments that include word processing drafts, games, and email correspondence.	Incomplete, fragmentary interactive computer games on floppy disks.
Access methods	Tiered and multiple: emulation, database, finding aid	Tiered and multiple: emulation, database, finding aid

Figure 8: Salman Rushdie and Timothy Leary comparisons

After a brief introduction, the following chapters will provide a descriptive summary across each of these categories for Salman Rushdie, followed by Timothy Leary, followed by analysis.

Despite their many manifest differences, Salman Rushdie and Timothy Leary share certain affinities, especially given the sheer range and riskiness of their public image. Their political stances on free speech, religion, or, in Leary's case, the free use of controlled substances both disrupted and endangered their lives for lengthy periods of time, while also permanently placing both in the public spotlight. While beginning his public life at Harvard University as a lecturer in psychology, Leary was later in and out of jail, or on the run, for his use and endorsement of narcotics. Throughout his multi-state, multi-country evasion, he nevertheless kept his personal effects intact, conscientiously saving and annotating important artifacts for the future.[105] His personal collections have been managed by a number of dedicated archivists, including Michael Horowtiz and, later, his estate Futique Trust, an outgrowth of his software company, Futique. Similarly, Rushdie the man of letters won accolades for his fictional work, but became better

[105] Robert Greenfield, *Timothy Leary: A Biography* (Orlando: Harcourt Press, 2006). 547.

known politically as an outspoken critic of Islam. His work forced him underground to avoid the very real threat of death. Both Leary and Rushdie had frequent encounters with law enforcement. The controversial values each espoused inspired passionate followers as well as detractors.

When the threat of death or jail was lifted, both Leary and Rushdie lived very public, high profile lives, frequently speaking on important issues, such as personal liberties, as well attending galas and parties with their many celebrity friends. During their respective public periods, they each released autobiographies. Leary's came before the donation of his archive, and suffers from major inaccuracies.[106] Rushdie, meanwhile, has credited the intellectual arrangement of his personal papers as benefiting his memoir *Joseph Anton* (2011).[107] Both Rushdie and Leary cultivated their reputations conscientiously in their awareness of the value of their personal artifacts and documents as well as in their personal memoirs. Given their respective controversies, it is no surprise that Rushdie and Leary valued authoring their own definitive autobiographical accounts. Their works, their associations with various cultures and subcultures, and the popularity and prestige of their acquaintances and friends amounts to a fascinating cross-section not of their lives but their times. As such, their newly accessible personal "papers" present many possibilities for wide-ranging research. The form and content of their hitherto private born-digital "papers" will give enterprising biographers and scholar-critics across the disciplines a lot to go on.

The following section is structured to provide collection overview across major categories characteristic of any born digital personal collection: organizational roles, donor involvement, content, copyright and intellectual property issues, and access approaches. As summarized in

---

[106] Greenfield, *Timothy Leary*.547.

[107] Erika Farr, interview with Salman Rushdie, "Salman Rushdie Discusses Creativity and Digital Scholarship with Erika Farr," Emory University, 2006, https://www.youtube.com/watch?v=kmb1oQcRmkM.

figure 8, the two cases demonstrate opposite characteristics across categories representing the breadth of mixed, hybrid personal collections. The subsequent affected processing strategies for each is so different as to represent a spectrum of possible outcomes in digital forensics processing of hybrid personal collections.

2.2 Salman Rushdie's Personal Computers

> *Rushdie stated that he first used his computer as a sophisticated typewriter, but as time passed and technology allowed, he slowly began incorporating all aspects of his life into his computers*.[108] -Erika Farr, Digital Archivist

Salman Rushdie is an Indian-born British novelist best known for *The Satanic Verses* (1988) and the resultant fatwa, or "death sentence," issued by Ayatollah Ruhollah Khomeini, the Supreme Ruler of Iran, against him as well as his publishers and distributors. (The fatwa was lifted in 1998.) The ensuing violence forced him into hiding for over a decade. *The Satanic Verses*, meanwhile, was a literary sensation. It received the 1988 Whitbread Award, was a Booker Prize Finalist, and took the eminent 'Booker of Bookers' prize as the best Booker Prize for Fiction in its twenty-five history. In 2007, Rushdie was made a Knight Bachelor for his services to literature. In 2008 Rushdie was elected a Foreign Honorary Member of the American Academy of Arts and Letters. Rushdie also served a term as a PEN American Center President and, in 2005, founded the World Voices Festival.

Rushdie joined Emory University's English Department as Distinguished Writer in Residence in 2007.[109] In 2006, his collection was purchased for an undisclosed price.[110] In 2011, his five-year appointment was subsequently renewed and he was also named a University Distinguished Professor.

---

[108] Laura Carroll, Erika Farr, Peter Hornsby, Ben Ranker, "A Comprehensive Approach to Born-Digital Archives," *Archivaria* 72 (2011), 65.
http://www.nytimes.com/2010/03/16/books/16archive.html?pagewanted=all&_r=0

[109] "Born Digital Manuscripts: A Primer" 2010 Rare Books and Manuscripts Section Preconference, transcript, http://www.rbms.info/conferences/preconfdocs/2010/2010_Seminar_I_transcript.pdf.

[110] Erika Farr, interview with Salman Rushdie, "Salman Rushdie Discusses Creativity and Digital Scholarship with Erika Farr," Emory University, 2006, nytimes.com/2010/03/16/books/16archive.html?pagewanted=all&_r=0. It was enough to pay his mortgage apparently.

The collection in total is vast and expansive. It encompasses his notebooks, photographs, and letters. Most interesting for our purposes, the collection also features nearly every personal computer in his personal computing history at the height of the turn to personal computing. The first of his personal computers, a Performa, was made accessible via emulation in preparation for the February 2010 opening ceremony of his collection. Early discussions with Emory representatives did not include talk of his collection of personal computers in his closet, kept long after their purposeful use due to ever-present security threats.[111] Two of his computers are scheduled for access in August 2014, with the last and most recent 2006 computer scheduled for December 2014.[112]

Rushdie's papers comprise 106.25 linear feet (215 boxes) and 55 oversized papers, but also include born digital materials of financial files, family photographs, legal files, audiovisual materials, daily calendar, emails, downloaded web pages, and games, with the majority dating 1992-2006. These born digital materials are processed chronologically. They are divided into the four computers and hard drive with the oldest computer processed first. Of additional interest is Rushdie's fax machine.[113]

1. Organizational Support and Roles

Rushdie's personal computers represent nearly his entire life with personal computing. Making them accessible was an unprecedented undertaking and would require significant resources. Vice Provost Richard E. Luce provided top-level support for a cross-departmental

---

[111] Kimber Williams, "Rushdie: Digital Archive at Emory 'allowed me to write' memoire," Emory Report. March 14, 2012] http://news.emory.edu/stories/2012/03/er_rushdie_digital_archives_memoris/campus.html.
[112] Dorothy F. Waugh, email to author, March 21, 2014.
[113] "Salman Rushdie papers, 1947-2008," Finding aid at the Manuscript, Archives, and Rare Book Library of Emory University, http://findingaids.library.emory.edu/documents/rushdie1000/.

working group to prepare materials for the 2010 opening ceremony. University libraries hired and allocated Information Technology support, programmers, and a computer security expert to form a "Rushdie Born Digital Archives Working Group" (BoDar) which extended beyond the Manuscripts (MARLB) staff to include the university's Born Digital Initiatives and the Digital Systems Departments (three members each, including: Naomi Nelson, Susan Potts McDonald, and Laura Carroll, Library's DSD; and Erika Farr, Ben Ranker, and Peter Hornsby). With the broad support of the University, MARLB was able to avoid the major stumbling blocks of many archives: shortages in staff and the lack technical expertise required for these collections. They developed detailed processing manuals and documentation, meeting weekly to discuss issues as they arose. Key members in the Rushdie processing include Erika Farr (Digital Archivist), Laura Carroll (Processing Archivist), Peter Hornsby (software engineer with library's software development team), and Ben Ranker (senior engineer for the libraries' software development team). The BoDar team encompassed skill sets that extend far beyond traditional archiving to also include digital humanities research, computer programming, content modeling, and Apple tech support.

2. Donor Involvement

Rushdie was an interested and invested participant in the process, able to step in to clarify his methodologies, privacy needs, and the significance of items lacking documentation. Rushdie's willingness to discuss donor issues as they arose clarified many ambiguities. Without an involved donor, Emory would have not been able to comfortably create access to many materials. Clarifying redactions relied heavily on the trust he developed with the staff. Erika Farr has noted that she and Rushdie spent time together reading his emails and reviewing what would or would not be redacted. Significantly, Rushdie, as a living and active donor, probably redacted

and deleted some of his materials before donation. While Rushdie allowed for the dark archive to contain any retained deleted material, the archivists respected the donor's wishes.[114]

While technologies have moved forward considerably in the past several years to streamline the redaction process, the BoDAR collection's need to identify any restrictive information with absolute certainty necessitated that *all* documents were screened by the processing archivist manually.

3. Content

> It is no more possible to capture a complete computing environment than it is to transfer or acquire a complete paper-based archive… A computer is a working environment that contains tantalizing traces and reminders that any single machine is part of a much larger material and virtual network and has relationships with a variety of other computers, devices, and servers not transferred to the archives.[115]

Delineating the boundaries of "a collection" with networked, complex, and wide-ranging personal computers is another crucial step that digital forensics can help untangle. Rushdie was a prolific journal keeper, a letter-writer, and drafter, but with the shift in technologies he found his process streamlined. Computers allowed him to "focus more on the writing" itself.[116] However, his computers are rich with interlaced materials about his day-to-day life, his writing habits, and his editing process. Resuscitating his computers gave researchers the ability to see outside these document, to correlate them with other computer-centered activities such as his email correspondences or even the Mac sticky notes he relied on for notes to himself. The possibilities for understanding and mapping his day-to-day personal life with respect to his writing practices by studying his calendars, his emails, his drafts, and other available documentation are now newly available.

---

[114] "Born Digital Manuscripts: A Primer," 20.
[115] Kirschenbaum, "Digital Forensics and Born-Digital," 21.
[116] Caroll, "A Comprehensive Approach," 65.

The Rushdie born digital collections notably included full computing environments: four Apple computers (one not functional due to a spilled drink). With the exception of one external hard drive from his then current, in-use, computer, there was no other significant removable media. The born digital collection totaled over 18 gigabytes of data and over 40,000 files. Significantly, of these files, approximately 11,350 were user generated (12,205 MB). In addition to the recovered previously damaged computer, his earliest Macintosh also included backups from his previous computer as his "OLD MAC."[117] Rushdie's born digital materials, then, comprised of his last five computers as well as back-ups of his oldest pre-1992 computer. Significantly, the finding aid relegates all of Rushdie's born digital material as the last series, Series 11, without much further explanation. In several sections of the finding aid, there are notes indicating when materials are either closed to researchers or available digitally.

The devices included:

Personal Computer Type	**Years in Use**	**Date available to researchers**
Macintosh Performa Desktop computer, 5400/180 (includes backups to older computer)	1992-2002	Feb 2010
Macintosh PowerBook laptop, 5300c	2002-2006	Aug 2014
Macintosh PowerBook G3 laptop, (QT9250B5G03)	2002-2006	Aug 2014
Macintosh PowerBook G3 laptop, (QT9386CEEY8)	2002-2006	Dec 2014
Smart Disk FWFL60 FireLite 60GB 2.5" FireWire Portable hard drive	2006 (then current)	Dec 2014

[117] Caroll, "A Comprehensive Approach," 71.

Figure 9, Rushdie's Computers[118]

In its processing of the collection, BoDAR attempted to shape access needs around traditional scholarship with literary manuscripts; that is, to show, document, and illustrate his creative process with multi-pronged access of different aspects of Rushdie's "content." This not only include text, but a rich environment of digital materials that could studied in relation to one another. BoDAR, then, processed the collection to make some of the writing process today with its multiple drafts and edits in digital form accessible to researchers.

The majority of his 1992-2002 era computer's accessible files include drafts of fiction, including *East, West* (1994), *The Moor's Last Sigh* (1995), and *The Ground Beneath Her Feet* (1999), as well as non-fiction drafts for the subsequent publication *Step Across this Line*. The workstation for the computer also provides access to selected email correspondences.[119] Some of these drafts are also available in the analog portion of the collection. Unique or not, it's important to attempt to write and understand implications to the content as interpreted inclusively to include the older mac environment and the "information" gleaned from the approximation of that experience of a smaller screen size, slower refresh rate, different (to our eyes, dated) interface, older proprietary software programs such as ClarisWorks word processing and Eudora email. "Content," from a media context, would include these aspects of the collection as well. While his

---

[118] Laura Carroll. "Arranging and Describing Born-Digital Archives: The Salman Rushdie Papers at Emory University," (presentation at the Society of American Archivist Conference, Chicago, IL, August 2011, http://www.slideshare.net/lcarro2/saa-2011-carroll.

[119] "The digital archives of Salman Rushdie: Overview" (Emory University Manuscripts Archives and Rare Book Library, researcher handout), http://marbl.library.emory.edu/documents/digital-archives/rushdie-help-sheet.pdf.

Performa Desktop is housed in a plexiglass for observation as an artifact of the recent past, some of the content and experiences associated with it are also available.[120]

4. Intellectual Property, Copyright, and Privacy

"I didn't realize what I was doing"

"It does feel a bit like undressing in public.[121]"

The privacy concerns were significant, but Rushdie allowed for the creation of bit-accurate disk image of his computers for preservation. That is, Rushdie allowed Emory University access to not only initially copy files, but to also preserve wholly as a dark archive the complete disk images of all of his computers which included deleted content, cached internet content, and his user history. BoDAR limited access to these images, sequestering them only as preservation backups to work from. The majority of staff working on the collection processing and access, then, were only given access to the already heavily redacted materials.

Entrusting the preservation image poised special risks. Collection material could include names and locations of the people who harbored Rushdie during the fatwa years, possibly endangering them. While the fatwa was over, there were still major security threats that went well beyond the normal discussions of privacy with personal papers. Beyond redacting electronic text, these digital materials could include hidden embedded metadata with geo-locations. These concerns were, in fact, the reason Rushdie had never thrown out his computers over the years. "We wanted to be extremely careful with the collection," Ms. Nelson told us. "If this got out and got

---

[120] Devyana Saltzman, "Snooping Through Salman Rushdie's Computer," *The Atlantic*, April 21, 2010, http://www.theatlantic.com/entertainment/archive/2010/04/snooping-through-salman-rushdies-computer/39260/.
[121] Farr, "Interview with Salman."

posted somewhere, it really could have been disastrous."[122] The restrictions that developed were then both important enough, nuanced enough, and pervasive to all portions of the collection that the BoDAR group, in fact, developed a workflow that was informed and motivated by these very restrictions.

From the beginning, the BoDAR working group were going to restrict significant portions for standard privacy reasons that complied with already existing norms in analog collections as well as law: all legal and financial files are closed until after Rushdie's death; all papers relating to family are closed until the death of the family member or 70 years after the acquisition date. All of his journals written after 1989 are also restricted indefinitely. The need to redact sensitive material from correspondences proved too resource-intensive during the initial phase of processing; for that reason MARLB has restricted almost all correspondence from his first computer.

These restrictions helped to guide researcher access developments. Through 2009, the processing archivist sorted content into categories that corresponded to the previously done arrangement of content in the paper-based items, including manuscripts, photographs, and financial records. The finding aid directly informed their tools and born digital content sorting. The BoDAR team also needed to balance privacy restrictions and created four access categories that will be elaborated on in the Access section. The nuanced restrictions on materials, however, were one of a number of major time and labor intensive parts to this collection. [123]

Access Approaches

---

[122] Jennifer Howard, "Rushdie's Digital Shadow, or How to Create Electronic Archives," *Wired Campus blog*, The Chronicle for Higher Education, September 24, 2012, http://chronicle.com/blogs/wiredcampus/rushdies-digital-shadow-or-how-to-create-electronic-archives/39690.

[123] Farr, interview with Salman.

	Emulation	Database	Finding Aid
As is	Yes	Yes	Yes
Redaction:	No	Yes	No
Emulation Only	Yes	No links in database	No
Restricted	No	No	Maybe

Figure 10, Restriction levels, from Carroll "Arranging and Describing."[124]

Access Types

PDF database	Includes Eudora email, fax, and text. Only accessible on-site
Emulation	Emulation of computer in a non-networked workstation. Only accessible on-site
Finding Aid	Accessible online. Limited description of born digital content.

Figure 11, Access types

In order to better understand how digital forensics have made the collection at all possible for research access, I include a simplified workflow of the steps in processing the Rushdie papers. In contrast to other archives that may offer files or even databases of the textual documents, they ultimately decided to include full emulations of each of the computers themselves in addition to a searchable database of the content of the files as each computer was processed and made available. They committed to preserving the original working environment and file types. This was a major innovation. According to Erika Farr, the director of born digital initiatives at Emory's Robert Woodruff Library, there is "no other place in the world that is providing access through emulation to a digital archive.[125]" "Most are offering discrete files," Farr has said. "We're providing numerous points of access into Rushdie's digital archive."[126] Farr elaborated that BoDAR, from the beginning, aimed to create an "authentic researcher experiences" that would include these emulations. They approached access methods early in their decision-making process.

[124] Carroll, "Arranging and Describing,"5.
[125] During the long period of processing, Erika Farr has changed titles.
[126] Patricia Cohen, "Fending Off Digital Decay, Bit by Bit," *New York Times*, March 16, 2010, nytimes.com/2010/03/16/books/16archive.html?pagewanted=all.

Copyright issues, such as the different proprietary software programs the emulations depended on, prevented them from making the archive too open, but the working group was dedicated to making the [holistic] complete computer environments available.[127] Thus, visitors can log onto a computer that simulates Rushdie's computer, giving access to his desktop wallpaper, computer ″stickies,″ and work drafts. The emulated non-networked workstation allows access to all three access points: the traditional finding aid, the word-searchable database, and the emulated early personal computer. The online finding aid is restricted and does not clarify exactly this tripartite structure and indicate the wealth of born digital material accessible onsite on non-networked stations at MARLB.

Workflow

1. Create a detailed spreadsheet of all hardware. This includes information such as serial numbers, processing type, speed, data, photographs, etc…

2. Create a disk image of each hard drive.

Create a complete bit-by-bit, checksum verifications of all material on each hard drive.

Create a metadata spreadsheet.

3. Use the checksums to weed out duplicate back-ups.

4. Create an untouched silo of all original hardware and data used for recovery and duplication, copy and create a 2nd silo that is similarly dark and restricted for preservation work.

5. Sort and group each set of files by content types.

6. Emulate the oldest computer, a Mac Performa 5400 (used 1992-2002).

---

[127] Mary J. Loftus, "The Author's Desktop" *Emory Magazine*, Winter 2010, emory.edu/EMORY_MAGAZINE/2010/winter/authors.html.

Use SheepShaver software to evaluate each file, refresh the file, and then set security and control levels for each document.

The workflow, even radically simplified, may help to also indicate that the processing for the Rushdie computers is very time and labor intensive. It cannot and has not been duplicated by any other archive for those reasons.

Conclusions

Rushdie's first computer, along with many of the analog materials in the collection, inaugurated the 2010 opening exhibit. The Rushdie materials were among the top ten most requested collections in 2010.[128] It has become a hallmark collection not only given the nature of its subject, Salman Rushdie, but also because of the pioneering nature of the collection. Since its opening, the collection has been subject of numerous publications. It has been featured in *the New York Times* and mentioned in numerous subsequent research on access issues to complex media. Since then, however, there are still very few accessible complex born digital personal collections and certainly none that require as much extensive individualized attention as the files of Salman Rushdie. One of the major questions posed by the still-incomplete processing of the Salmon Rushdie collection is whether or not we should assume that such intensive work is the goal of major archives. Are we looking more and more at servicing a tiny percentage of the backlog of complex legacy material?[129] Is there a way to simplify these approaches in a less resource intensive way, modeling MARLB's own recent adaptions as they process other collections?

---

[128] Salman Rushdie papers, 1947-2008," Finding aid, 126.

[129] On the other hand, complex environments such as these provide layers of context and have consequence far beyond legacy born digital media to include databases, scientific data, and networked machines. Digital forensics cases such as these offer possible solutions to non-traditional, dynamic "objects," such as computer games in which there is no singular object for preservation. [16].

Years later, these questions and concerns are also posed by the staff. MARLB has received several more born digital records of writers, but in the cases that have followed there has not been anywhere near the same need to repeat this level of complexity. In some ways, the complexities of Rushdie are unique: nuanced and detailed privacy levels and multiple full-computing environments dating back into the 1980s. MARLB has found that the time, coordination, resources, and staff marshaled for the one computer cannot readily be imitated, even by Emory. For the remainder of the collection's personal computers, fewer people are involved. It has necessarily become much less time-intensive.[130] Practices such as the use of spreadsheets have been tailored to avoid the high chance of human error and inessential work. In a recent update on workflows, Erika Farr discussed some of the complications in finishing the project. The team as originally assembled is not still intact. The collection is no longer as high on the institutional agenda. There are other collections to process. Their newer archivist has just finished sorting through the 2nd and 3rd computers. She is on the first sort of the last computer, the hard drive Rushdie had deposited of his then in-use personal computer. The process has been vastly simplified. Mapping the fields onto the digital repository was cumbersome, time-consuming, and prone to error. Now the processing mimics the paper-based workflow in which files are sorted into directories created and divided by restrictions and series and subseries. This simplified the metadata issues as well; metadata is batched by groupings rather than determined on a document level. However, years later the collections is still in-progress.[131]

Through the word-search function of his digital material in PDF format, the expectation is that researchers can delve into quantifications of word usage and better find any themes, patterns, and parallels across different aspects of Rushdie's life and writing practices. Through the

[130] Erika Farr conversation with author.
[131] Ibid.

combination of its charismatic and controversial subject, its innovative and well-publicized innovations, and the extensive overlapping disciplinary research, the Rushdie papers are already one of the most popular items in the collection.[132]

[132] Caroll, "A Comprehensive Approach," 87.

Timothy Leary's Games

Turn on, boot up, jack in.[133]

The PC is the LSD of the 1990s - Timothy Leary[134]

Timothy Francis Leary, Jr. (1920-1996) is primarily known as "a psychologist and writer […] an advocate for the use of psychedelic drugs and a countercultural icon."[135] While a seminal pop cultural icon of the 1960s, Leary was a chameleon, and his life took many unpredictable turns. Freewheeling and experimental, he associated with the Beats, the Weather Underground, Eldridge Cleaver, John Lennon and Yoko Ono, and Hollywood celebrities such as Susan Sarandon, his godchild Winona Ryder, and technology pioneer Joi Ito.

Leary lectured at Harvard University's Center for Personality Research, and with Richard Alpert/Ram Dass, researched LSD's therapeutic effects with subjects like Allen Ginsberg, until his 1963 dismissal. Leary became a major advocate for drugs usage ("Turn on, tune in, and drop out"), inspiring then President Richard Nixon to call him, "the most dangerous man in the United States."[136]

The authorities systematically targeted him for his views and his behavior; he was subsequently tried and incarcerated for possession of marijuana with a sentence of 20 years.[137]

---

[133] Arno Ruthofer, "Think for Yourself; Question Authority: the development of Timothy Leary's theories and his impact on the psychedelic-cybernetic counterculture," (master's thesis, 1997), introduction, archive.org/web/20061023201632/http://www.geocities.com/arno_3/intro/ or http://issuu.com/scottjenson/docs/arno_ruthofer_-_think_yourself__que.

[134] Philip Elmer-Dewitt, "Cyberpunk: Virtual Sex, Smart Drugs, and Synthetic Rock 'N Roll," Time Magazine 141:6, (February 1993). https://archive.org/stream/cyberpunkvirtual00time/cyberpunkvirtual00time_djvu.txt.

[135] "Timothy Leary papers, 1910-2009," Finding aid at the Archives and Manuscripts Division of the New York Public Library, http://archives.nypl.org/mss/18400.

[136] Laura Mansnerus, "Timothy Leary, Pied Piper of Psychedelic 60s, Dies at 75," *The New York Times* June 1, 1996, nytimes.com/1996/06/01/us/timothy-leary-pied-piper-of-psychedelic-60-s-dies-at-75.html.

[137] The title of Timothy Leary's spoken word album (1967) and eponymous documentary (1967), dir by Robin S. Clark.

Championed by outraged liberals, he was able to arrange a jail-break through connections with the student political-terrorist splinter group, The Weather Underground, and the Black Panthers.

After his re-capture and eventual release from prison, he re-fashioned himself many times, working as an actor, a public speaker, a software developer and a writer. In 1982, Leary founded Futique, Inc. to focus on computer and technology projects. His interest in the area of personal computing led to his collaboration with software companies such as Interplay, Electronic Arts, and The XOR Corporation. While not necessary a technologically adept person himself, Leary's charisma allowed him to gain the support of many artists and celebrities for his many ventures.[138]

In 2011, the New York Public Library purchased the collection from the Futique Trust for $900,000 with a portion of the purchase price reportedly funding the collection's processing.[139] In 2013, the Leary Papers were opened to an excited public especially interested in his personal digital media. Previous to the collection's processing, his born digital media was thought long lost; no provisions had been made for its accessioning in the purchase agreement. This part of Leary's life, his fascination with "cyberspace," futurism, and software development is notably missing in published research. Greenfield's definitive biography of Leary barely mentions his software development period. Leary, a consummate collaborator and visionary, worked with seminal figures in technology and design. In a later section devoted to "Content," I will limit itself to addressing Leary's born digital collection and reconstruct his various attempts to create and launch software programs. While the Leary collection also contained of "electronic records" of text-based documents that required significant work before being made accessible through a

---

[138] Memos dated March 20, 1986, March 23, 1986, box 217, folder 21, Timothy Leary papers, MssCol 18400, Manuscripts and Archives Division, The New York Public Library.

[139] Patricia Cohen, "New York Public Library Buys Timothy Leary's Papers," The New York Times (New York, NY), June 15, 2011, nytimes.com/2011/06/16/books/new-york-public-library-buys-timothy-learys-papers.html?pagewanted=all.

database, in this thesis I will focus primarily on the digital forensics necessitated by Leary's software and "games," following the same structure as in the Rushdie case study.[140]

The Leary archive contains 263.78 linear feet (575 boxes, 40 sound recordings) and 1.78 gigabytes of data (4,043 computer files, 56 disk images).[141] Jennifer Ulrich processed the collection. Digital archivist Donald Mennerich and intern Alison Rhonemus are credited with the digital forensics.[142] Publicity from Wired Magazine and other publications announced the availability of the Leary archive to the public on September 18, 2013.[143]

His analog archival material holds records of meetings, plans, and schematics but many of his projects remained incomplete. An additional complication was that the intellectual arrangement involved considerable guesswork and interpretation. The disks were not meaningfully arranged, but were found "fugitive" in a box. The accompanying analog documentation was also unclear.[144]

1. Organizational Support and Roles

The Manuscript and Archive Division of the New York Public Library worked independently of other divisions of the library and did not, for example, have significant extra support from the information technologies department. The majority of the work behind the Leary papers involved the Digital Archivist, Donald Mennerich; the Processing Archivist, Laura

---

[140] "Timothy Leary papers, 1910-2009," Finding aid.

[141] Ibid.

[142] Don Mennerich has since gone to New York University to become the first appointed Digital Archivist of the libraries. Alison Rhonemus is currently employed as a part-time staff member handling the digital forensics station.

[143] Greg Miller, "Timothy Leary's Transformation from Scientist to Psychedelic Celebrity," Wired, October 1, 2013 wired.com/2013/10/timothy-leary-archives/. wired.com/2013/10/timothy-leary-archives/. Also, James Joiner, "Get Your Hands on Timothy Leary's Stash," *Esquire blog*, September 18, 2013, http://www.esquire.com/blogs/news/timothy-leary-papers-released-to-public.

[144] Chris Person, "The Lost Computer Games of Timothy Leary,"Gizmodo.com, November 6, 2013. Video interview, gizmodo.com.au/2013/11/the-lost-computer-games-of-timothy-leary-2/.

Uhlrich; and intern Alison Rhonemous. A workflow was created and streamlined with this project, but the departmental support needed to maintain access and support to some portions of the archive were not re-assigned when key staff members departed.

2. Donor Involvement

There was no donor involvement in the born digital processing. Timothy Leary passed away in 1996. Leary had prepared for the sale of his personal archive while alive, morphing his software company "Futique" to his estate "Futique Trust." Leary was highly invested in keeping his archive intact during his life. In 1970, he hired Michael Horowitz, who served as his personal archivist for several decades. The Leary estate, however, did not significantly involve itself with the processing.[145]

3. Content

Leary's software is "fugitive." Approximately 375 disks were found by the archivists long after the agreement had been signed and the materials had been already transferred to NYPL for arrangement and description.[146] The born digital aspects of the collection were composed entirely of disks. The disks containing software that ran on IBM PC, but there were also a number of Commodore Amiga pieces of early electronic artwork created by collaborator, artist Keith Haring. In addition to these, there was also a single 44 megabyte Syquest tape with unknown content on it that was not, ultimately made accessible due to obsolescence.[147] While the Digital

---

[145] Donald Mennerich, conversations with author, March 21, 2014.
[146] Jennifer Schuessler, "Timothy Leary Video Games Unearthed," The New York Times, September 27, 2013, gizmodo.com.au/2013/11/the-lost-computer-games-of-timothy-leary-2/.
[147] Trevor Owen. "Born Digital Archival Material: An Interview with Don Mennerich."http://blogs.loc.gov/digitalpreservation/2013/04/born-digital-archival-materials-at-nypl-an-interview-with-donald-mennerich/.

Archivist spent significant resources trying to find a way to gain access to the content on the Syquest tape, none were ultimately successful. Also, while Leary's personal computers were originally included in the Futique Trust, they were not included in the sale with the rest of the collection.[148] However, Leary's Nintendo Power Glove, an example of his early adoption of gaming and cyber culture, was part of the collection.[149]

Without clear documentation to work from, the archivists arranged the media after analysis in an emulator. The archivists chose to interleave the born digital media throughout the finding aid (rather than, for example, creating a separate series for the born digital media). Media was divided into two broad categories of primarily text ("electronic records" or "ER") and complex media ("disk images" or "DI").[150] This meant that *Neuromancer* disk images from multiple files could be combined under the same "DI." Different media with the same content-*Neuromancer*-were arranged together in the finding aid, regardless of whether in text form, analog form, or disk image. By creating this more comprehensive finding aid, researchers were able to glean some of what was available in electronic form online offsite.[151]

Some of the content on the disks had analog counterparts in the form of dot-matrix-printouts. However, much of the born digital content was fragmentary, incomplete, or difficult to "run" on contemporary computers. The difficulties presented by these materials will be discussed more thoroughly with several examples at the end of this section.

---

[148] Daniel Terdiman. "Timothy Leary's archives: Bridge from '60s to '90s" 7 Feb 2009. CNET. cnet.com/news/timothy-learys-archives-bridge-from-60s-to-90s/

[149] "We just had to share this with you, "NYPL Wire- The New York Public Library Tumblr, November 28, 2012, http://nypl.tumblr.com/post/36741758673/we-just-had-to-share-this-with-you-while-taking-a.

[150] The term "electronic record" is ambiguous today, especially when there are other digital items excluded from this category. NARA defined the term as "any information that is recorded in a form that only a computer can process and that satisfies the definition of a Federal record per the *Federal Records Act." From "Context for Electronic Records Management," National Archives, archives.gov/records-mgmt/initiatives/context-for-erm.html.*

[151] "Timothy Leary papers, 1910-2009," Finding aid.

Timothy Leary's Software Games

You call yourself a philosopher, a reformer. Fine. But the key to your work is advertising ... You must use the most current tactics for arousing consumer interest.[152]- Marshal McLuhan

Despite his own technical limitations, Leary was an early proponent of personal computing. As psychologist Humphrey Osmond said, Leary is "someone who lives in an almost totally hypothetical future."[153] Before Futique Trust became his estate, Futique was initially oriented as a software development company, Futique, Inc. (1982).

In Robert Greenfield's biography, scant text is devoted to Leary's software and futurist enterprises in the 1980s. Greenfield did note, however, Leary's attempts to learn about computing from a young UCLA film archivist Jeff Shtiel in order to develop interactive software for the film industry. Their partnership resulted in the program *Box Office Prediction Scale* (BOPS), also referred to as the *Shtiel Film Aptitude test*.[154] The two devised a computer program to evaluate a film protagonist's strength, correlated with box office success or failure. This software program was later incorporated in Leary's first and only marketed software program, *Mind Mirror*.

Mind Mirror (1985)

Computers can be used as intelligence appliances (i.e.) for mental fitness, self-improvement.[155]
--Timothy Leary

---

[152] The American Conservative, theamericanconservative.com/articles/the-acid-gurus-long-strange-trip/

[153] Luc Sante, "The Nutty Professor," The New York Times, June 25, 2006, nytimes.com/2006/06/25/books/review/25sante.html?pagewanted=all

[154] Robert Greenfield, Timothy Leary, 551

[155] "Computer as Mind Appliance." Mind Mirror Video Proposal, 4 Feb 1988, box 217, folder 21, Timothy Leary Papers. The New York Public Library Archives and Manuscripts. Memo written as if a proto-internet device for communication had already been developed: "the mind mirror appliance makes it possible for everyone working on the film to Tele-Thought (screen) their ideas about characters, actors, script development. Futique is prepared to set up a Tele-Thought Communication system so that at every stage of production everyone's ideas about the film can be communicated clearly."

Mind Mirror was created in collaboration with programmers Peter Van den Beemt and Bob Dietz and released by Electronic Arts in 1985 for use with Commodore 64, Atari XL, Apple II, and MS-Dos. It ultimately sold 65,000 copies.[156] As its name suggests, *Mind Mirror* explores themes of interpersonal development and has been described as a "digital reinterpretation of his 1950 doctoral dissertation, 'The Social Dimensions of Personality.'"[157] Players create and evaluate personalities and, as a result, are able to map their personalities onto a 360-degree spectrum of personality types, which Leary called the "Interpersonal Circle" or "Interpersonal Circumplex." This Circle or Circumplex is divided into four types of personalities, further divided into sixteen types, as previously elaborated in Leary's "Interpersonal Diagnosis of Personality."[158] More recently, the rights to the program were sold by to the original developers for development and updating for modern computing systems and a new version has been made available through Facebook.[159]

Mind Mirror was the only commercial release made. While Futique continued to research other games, their projects are less clearly charted beyond *Mind Mirror*. Leary frequently had multiple projects in development and used similar titles and explored similar themes. Overall, however, he was motivated by his embrace of the future's possibilities with the developments in computing. Other than the never completed *Neuromancer* project, the collection includes innumerable files and folders documenting many closely related projects such as *Mind Movie*, *Mind Adventurer, Mind Play, Intercom,* and *Interscreen,* among others. In all of these projects,

---

[156] Timothy J. Seppala, "Timothy Leary Invented Video Games that Found in New York Public Library," Engadget, October 2013, http://www.engadget.com/2013/09/30/timothy-leary-video-game-archive-drugs-are-bad-mmkay/
[157] "Timothy Leary," Wikipedia, http://en.wikipedia.org/wiki/Timothy_Leary.
[158] "Mind Mirror," Timothy Leary Archives, timothylearyarchives.org/mind-mirror/.
[159] "Mind Mirror," www.mindmirror.com.

the "games" that Leary tried to create were highly informed by his lifelong interest in transpersonal games that would enable personal development through understanding one's personality type.[160]

Neuromancer

William Gibson's award-winning dystopic cyberpunk novel, *Neuromancer* (1984) inspired Leary to take on an ultimately never completed software project. *Neuromancer* was one of the first science fiction novels to explore cultural transformation due to the information age, inventing the term "cyberspace."[161] Gibson's proto-internet introduced millions of readers to the concept of a global network of computers and of virtual reality, or a habitable "place" on the web before these concepts were in widespread development and use. His work's iconography has been elaborated in popular films such as *The Matrix* (a term he also originated). The ideas of *Neuromancer* may have also directly influenced technological developments.

Timothy Leary purchased the rights to create an equally ambitious game design, engaging many of high-profile collaborators on the project that included Devo, Keith Haring, Helmut Newton, Grace Jones, David Byrne, and William S. Burroughs, and other contemporary luminaries he had befriended. His work spanned a number of years, but ultimately his versions were never completed.[162] Leary had brought the project to Interplay for development and they eventually released a *Neuromancer*, crediting Leary as a producer.[163] The Interplay version of *Neuromancer* was eventually released on AMIGA OCS 1989, retaining Devo's contribution, a soundtrack based

---

[160] Timothy Leary papers, box 229.
[161] David Wallace-Wells, "William Gibson, the Art of Fiction No.211," The Paris Review, theparisreview.org/interviews/6089/the-art-of-fiction-no-211-william-gibson.
[162] Neuromancer promotional video, https://www.youtube.com/watch?v=kye37wiui2A.
[163] Schuessler, "Timothy Leary Video Games."

on "Some things Never Change." In 1996, the year of Leary's death, *Computer Gaming World* awarded it the "Adventure Game of the Year."[164]

Leary initially imagined that *Neuromancer* would be like a choose-your-own-adventure-style "movie" in which each protagonist would explore a number of possible courses possible within the movie outline of 8 acts, 32 scenes, and 128 takes.[165] The archive contained diagrams document the effect of each choice possible in N*euromancer* (The choose-your-own-adventure format recurred across many of the games I accessed as a researcher.). Each scene was to have its own graphical art accompaniment, and each of the ten roles were to have four graphics. Thus there are 168 illustrations, but the software programs do not include animations as we understand them. The graphics were still images that faded or flashed different colors and geometries, and are characteristic of early computer animation.[166]

4. Copyright and Intellectual Property Issues

While the copyright for material created by Leary is maintained by the Futique Trust, the obstacles associated with intellectual and copyright issues with born digital collection materials are uniquely not present with the Leary collection. There has not been any substantial need to redact work or even keep some materials "dark" and inaccessible. Nor has there been any issue with third-party software. Leary's created programs do not rely on third-party software programs, but are executed code itself. Leary's textual work was migrated off of any proprietary program dependence with an accessible database. Leary did not participate with emailing. The

---

164 "Neuromancer," Wikipedia, http://en.wikipedia.org/wiki/Neuromancer_%28video_game%29/.
165 Timothy Leary papers, box 299, folder 9.
166 Timothy Leary papers, box 231, folder 1.

text-based documentation made available has all been migrated into a text-searchable database independent of any closed and commercial software system. The work accessed did not include anything of a very private nature; the disks imaged were primarily related to work projects.[167]

5. Access Approaches

Full-text Database	Accessible onsite. Text-based electronic records.
Emulation	Emulation of DOS environment for disk image access in a non-networked workstation. Only accessible on-site.
Finding Aid	Accessible online. Significant born digital content noted.

Figure 12: Leary Access Types

Unlike Rushdie, there is no sense of restriction levels playing a role in determining access types and levels. Access to all electronic records are available only on-site at a non-networked computer. The Leary archive was made available at a single computer that also was the primary access point for the other accessible born digital collections. Throughout my research visits, the Reference Archivist Tal Nadan primarily worked with me to answer questions (e.g., "Why won't this disk image load?"). I was able to experience, first hand, some of the ways in which new researcher access may develop with complex born digital materials. The norms regarding what and how much to support researchers are not established yet. Is this simply akin to researchers accessing a foreign-language work, for example, or something that requires more support from archives? At my initial visit, I was not given any written guidelines for the emulator DosBox. Nadan made sure that for my next visit an edited example of the tutorial would be made available for my use.

[167] Donald Mennerich, conversation with author, March 21, 2014.

The materials I looked at were primarily made available through a DosBox emulation. In addition to DosBox, SheepShaver and Quickview Pro also installed to provide access to other born digital material. Each of these three programs are well-known crowd-sourced emulators that provide broad support to recreated older environments.

Hardware emulation has existed for a long time in computer science as well as gaming communities. Popular emulators include Basilisk II, SheepShaver, QEMU, MESS, and MAME. Finding the appropriate emulator involves matching the processor architecture (ex: Apple would include Motorola M68000, PowerPC, or Intel x86) and the Operating System in use.[168] DOSBox is an open source emulator of Intel x86 PCs. It can run on modern personal computers and operating systems and allows access to many MS DOS era games, as well. SheepShaver is an open source emulator that allows one to run classic Mac OS applications. While originally a commercial application (1998), it was re-released in 2002 as an Open Source piece of software.[169] Both these emulators are widely used in the video game world. Lastly, the third piece of essential software is QuickView Pro, a commercially available software that allows viewing of many file format types and their metadata, marketed primarily to the legal forensics world. While the Avantstar website boasts that it will allow one to "view virtually any electronically stored information,"[170] this is not quite the case. Marketed primarily towards law and computer forensics, QuickView Pro is yet another example of an adapted tool from a digital forensics workflow that does not support the very old file formats included in many collections. As a contemporary piece of software, it was unable to open most of the components of the Leary disk

---

[168] Ian Welch, Nicklas Rehfeld, Euan Cochrane, " A Practical Apprach to System Preservation Workflows" http://www.researchgate.net/publication/235956241_A_Practical_Approach_to_System_Preservation_Workflows

[169] SheepShavher website, accessed August 20, 2014, http://sheepshaver.cebix.net/#faq

[170] Avant Star Quick View Pro http://www.avantstar.com/metro/home/Products/QuickViewPlusProfessionalEdition/Info/Specifications

images. Many of the individual files of each disk image, which included, for example, SYS, BAK, MP, SQZ, MP, and VIDPOP, are simply components to executable coding on the same disk image.[171]

Workflow[172]

1. Image the floppy disks

   To image the disks, they used Kryoflux F250.

   If cannot read sector format, image streams for later work

   Floppy disks are imaged as –Raw (DD)

2. Extract metadata from image's file system

   For FAT12 format, use Dfxml

   For other formats, create file hash from FTK Imager, csv file.

   Forensic Toolkit analysis and previewing, if file format allows.

3. Photograph the object (store in archival boxes).

4. Make a backup of the image, metadata and photographs to the cloud,

5. Transfer the image, metadata and photographs into Archivematica. [173]

6. Emulate. Use Apache Tika, SheepShaver, and Basilisk.

Conclusions

While I include images, diagrams, and directories from the digital and analog collection that pertain to *Neuromancer*, they do not fully inform us about the software programs created.

---

[171] Timothy Leary papers, disk image I5, disk image 13.
[172] Emails to author by Alison Rhonemus. February 24, 2014.
[173] conversation with Don Mennerich

Each of these documents support the notion that access to complex media needs to be multiple and multi-faceted.

While I explored the digital and analog components of the project for further investigation, it became clearer to me that understanding Leary's vision for *Neuromancer* went well beyond the scope of my initial aim to document the born digital collections portions and the problems and solutions. "Playing" Timothy Leary's games in a contemporary sense would involve an active re-interpretation of the incomplete games for a collaborative authoring with the embryonic ideas roughly sketched by Leary.[174] My experiences exploring both the analog and digital material of the Leary collection influence my thesis; I was forced to think more and more about the researcher's evolving role in accessing these collections. Not only are access types different, content types and technical competencies are also different. Thus, in my concluding chapter I devote a significant portion of text exploring how complex born digital collections such as Leary and Rushdie change the way researchers work.[175]

In the next section on access, I will explore how legal forensics workflows relate to those questions of access which problematize familiar archival principles. While disk imaging is something many institutions may be readily equipped to undertake, the problem of access is much more difficult. As a field, we will need to understand these implications before moving forward.

---

[174] "Playing Timothy Leary's Lost Games." Kotaku. November 5, 2013. You Tube. http://www.engadget.com/2013/09/30/timothy-leary-video-game-archive-drugs-are-bad-mmkay/

[175] If I have time, in chapter 3 I also think it would be good to map out implied essential characteristics and behaviors for pieces and problematize these frameworks.

Chapter 3: Access

Lack of use is one of the most serious threats to long-term preservation of digital collections.
--Christopher Lee and Kal Woods[176]

While archives are putting enormous effort in instigating digital forensics, it is important to extend the framework from accession all the way to access. This chapter first frames the issue through 1. Researcher studies 2. Discussing the different access types now available 3. Illustrating these access types through case studies. In the final section, we will discuss how the new affordances of digital forensics challenge and test foundational archival principles.

Researcher access to born-digital collections is relatively unexplored area of study. While few archives currently have any accessible complex born digital collections, it is important to begin posing these questions to ensure that archives are selecting and interpreting work in a way relevant to researchers. The interpretive work involved in access requires this foresight. We may better know the answers to some of the questions in the next 5-10 years as researchers become more aware of the possibilities of born digital research.

Valuations by different stakeholders affect processing approaches to born digital collections. Researcher interest in manuscript marginalia, frame markings on celluloid film, and newspaper advertisements, for example, have all influenced preservation processes. In other words, the screen size or time date stamp may serve as "mere" context for a traditional literary scholar; but for a digital humanist, it may verge on the very content itself. One question with important implications is: how have researchers used highly publicized born digital collections? How are these materials interpreted and used in the burgeoning area of Digital

[176] Christopher Lee and Kam Woods, "Automated Redaction of Private and Personal Data in Collections: Towards Responsible Stewardship of Digital Heritage," The Memory of the World in the Digital Age: Digitization and Preservation, 2012. Vancouver, BC, http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/VC_Lee_Woods_26_D_1600.pdf.

Humanities?[177] Emory University's MARLB analyzed researcher use of their first emulation of Salman Rushdie's Performa personal computer. Through their researcher feedback, they realized that there was confusion over some of the redactions, for example, in the emulated materials. Researchers would click to empty folders, unaware that the contents had been redacted from the emulated computer. After discovering this, redactions in the emulation were noted unambiguously.[178]

These questions are important to understand how to better provide access to complex material, but also to better advocate for their preservation and for robust workflows and scalar methods to be implemented with institutional-wide support. Cornell University's "Preservation & Access Framework for Digital Art Objects," funded by the National Endowment for the Humanities, created a survey for educations, curators, artists, and researchers to draw out particularities in each of these user group's practices to, in turn, better inform their developing framework for preservation.[179] As principle investigator Oya Rieger noted, "Reproduction of an artwork's digital files does not always insure preservation of its most important cultural content. It is essential that we anticipate the needs of future researchers and acknowledge the core experiences that need to be captured to preserve these artifacts."[180] With such variance in core experiences, Cornell wanted to make sure to take into account all stakeholders. In their first year of funding, Cornell sent out surveys, receiving approximately 100 − 120 responses with a completion rate of approximately 90%. Access questions ranged from how important was original hardware

177 Matthew Kirschenbaum, Born-Digital Forensics seminar class discussion, University of Maryland, College Park, MD, August 8, 2014.
178 Caroll, "A Comprehensive Approach," 87. They annotated any deletions from the emulations because of noted researcher confusion at empty folders.
179 Digital Media Art Preservation, CAA News, College Art Association. CAA, last accessed August 20, 2014, http://www.collegeart.org/news/2014/01/24/digital-media-art-preservation-project-questionnaire/
180 "Humanities grant helps library preserve digital art." Cornell Chronicle. http://news.cornell.edu/stories/2013/02/humanities-grant-helps-library-preserve-digital-art.

and access to code, to questions on screen resolution and size. Some stakeholders are disinterested in or uninformed regarding these distinctions; others find fidelity to the original environment and experience paramount. Results from the survey have just been released, but one of the resounding takeaways is the need to account for stakeholder interests, especially those of the creators and artists of the media.[181]

The Harry Ransom Center at the University of Texas acquired Michael Joyce's papers in 2006, and it stands as one of the very early pioneering collections processed when archives were grappling with electronic hypertext. The Ransom Center offers access to its digital materials on a case-by-case basis. In the last two years, three patrons have requested access to the Michael Joyce materials and the Center has been able to accommodate all of them. Two patrons used the Arnold Wesker and Michael Joyce materials via DSpace in the reading room, the third patron accessed disks from Terrence McNally's collection from a reading room laptop.[182] This seems a scant number. Archives are under significant financial pressures and complex media processing is resource intensive. Archives are not immune to evaluative quantitative metrics. Following analog archival conventions, benchmarks for patron access and resulting publication may be low. Nevertheless, where are the researchers and the work?

One observation by Matthew Kirschenbaum is simply that no groundbreaking work has emerged to take full advantage of these new media and access types.[183] Traditional scholars may not value using an emulator; they may simply want to read the text without noting its remediation.

---

181 "Interactive Digital Media Art Survey: Key Findings and Observations," Cornell University Digital Scholarship and Preservation Services, July 30, 2014, http://blogs.cornell.edu/dsps/2014/07/30/interactive-digital-media-art-survey-key-findings-and-observations/.

182 Matthew Kirschenbaum, Erika Farr, Kari Kraus, Naomi Nelson, Catherine Peters, Gabriela Redwine, Doug Reside, "Approaches to Managing and Collecting Born-Digital Literary Materials for Scholarly Use," Offices of Digital Humanities, National Endowment for the Humanities, 2009, 11.

183 Kirschenbaum, Born-Digital Forensics seminar. Of course he is discounting his own pathbreaking work in this intersection.

Other possible explanations include that researchers may also find accessing older complex media burdensome. The case studies each necessitate navigating of older operating systems. The types of scholars interested in media theory and media archaeology have traditionally worked with much older collection materials.

Nevertheless, even in the cases analyzed we can see that a number of access types are now possible due to digital forensics. The tools of digital forensics supply new ways to manipulate and to access material that were previously impossible. Making collection material accessible involves answering questions such as: should items be accessed on a file level or system level? What or how much change is acceptable? Should deleted material be made accessible? Can disk images be treated as access copies for researchers to take home as well as preservation copies? While some of these questions have been cursorily addressed, this next section will illustrate interpreting access types with examples for previous cases as well as some new ones.

Access Types:

In order to account for different security issues as well as issues of "authentic" experiences and "essential qualities," complex born digital collections require multiple access points. By that, I mean that archives must re-think access with complex digital. With complex media, accessing a migrated file may be insufficient. The complexities in a computing environment or software program require multiple attempts in order to replicate authenticity in the aggregate. As already seen with digital forensics toolsets and workflows, there is no clear singular solution. Reenacting the original experience is impossible; archives may need to do their best with several approximations that approach the question of content and essential properties from different

angles. The following list encompasses many possible access points, but as we saw in the Salman Rushdie and Timothy Leary case study, multiple access types are preferable.

- **On-site emulated environment of older system**

This is necessary for non-contemporaneous computing systems, software programs, video games, and any other complex born digital material with behaviors and characteristics that would be lost (or inefficiently preserved) with migration.

ex: Emory University, Salman Rushdie computer SheepShaver-based emulation of Mac Classic.

- **On-site disk image mounted in emulated environment**

When disk images are access points, but the computing environment needed to access the disk image is non-contemporaneous.

ex: New York Public Library, Timothy Leary disk images and electronic records mounted into 1980s era computing environments.

- **On-site word searchable database**

Text is normalized for non-linear database access with word-searching capability. Access to these materials, however, would still require significant work.

ex: New York Public Library's Gay Men's Health and September 11th collection files.

- **On-site access of normalized files in contemporary computing system**

This is the most common born digital access available because of its relative simplicity in comparison to different types of emulation. Text may be normalized to PDF, images may be normalized to JPEG, and music may be normalized to MPEG-3.

ex: Emory University's MARLB has done this for their Alice Walker born-digital materials.[184]

- **Online access to records**

This is not possible for most collections because of intellectual property and privacy considerations. The government records were all open to the public already and did not include any private materials. The Rorty collections is screened; permission must be granted through an online application. The contents of the Rorty collection are also not searchable through Google. BitCurator Access is focusing on creating online accessible rich metadata and even access.

ex: University of Irvine, Richard Rorty Collection text material made available online.[185]

ex: University of Indiana, Bloomington, Government Records[186]

- **Onsite access to cross-collection searching**

"Cross-drive" analysis is something that digital forensics will make more possible.[187] Stanford University Libraries allowed researcher to use FTK software to search across collections.[188] That is, instead of scholars searching from box to box, within each collection, this digitizing and text normalization allows researchers to search granularly across collections.[189]

---

184 "Alice Walker, 1944 - ," Finding aid at the Manuscript, Archives, and Rare Book Library of Emory University, Series 13, http://findingaids.library.emory.edu/documents/walker1061/series13/?keywords=alice+walker.

185 Richard Rorty born-digital files, University of California Libraries, http://ucispace.lib.uci.edu/handle/10575/7.

186 Virtual CD-ROM/Floppy Disk Library, IU Digital Library Program, Indiana University Bloomington, Libraries, http://webapp1.dlib.indiana.edu/virtual_disk_library/index.cgi/4298169.

187 Simson Garfinkel, "Forensic Feature Extraction and Cross-Drive Analysis, *Digital Investigations* 3S 2006:71-81, http://www.dfrws.org/2006/proceedings/10-Garfinkel.pdf.

188 Lee, "From Bitstreams to Heritage," 31.

- **Online access to topic-modelled collections**

Stanford's work on topic modeling, now in beta, will give researcher the ability to create metrics of personal collections to, for example, see how often a word is used. Researchers will be able to do "sentiment analysis" searchers that analyze weighted words for different emotional states to analyze sentiment. As seen in the examples, researchers will also be able to visualize networks of relationships between email corresponds and create network visualizations.

Ex: Stanford University's Richard Creeley emails.[190]

While archives may have reputations of neutrality, a significant investment in interpretative processing happens before these items are made available in any way. These interpretations have lasting impact on how a collection item, whether disk image or game, is perceived by future researchers, especially as we move farther and farther away from original computing environments. These considerations do not take into account the initial collection prioritization. While digital forensics adoption will make more collections accessible, very few collections will get the same level of consideration as the Leary and Rushdie papers did.

[190] "EPADD: Email, Process, Access, Discovery, Delivery, beta site," Stanford University Libraries, Special Collections and University Archives, http://epadd.stanford.edu/muse/archives/.

Conclusions

With these developments, archivists with newly granular-level tools available through digital forensics now have the option to disregard traditional principles such as the fonds, original order, measurement types (disk? File? MB?), (EAD), and finding aids. The implications of new access types will be explored further in the following conclusions.

The Fonds and Original Order

Respect des fonds mandates that the records of the creating entity not be mixed with those of a different entity. It prioritizes the archival records, "identifying the locus of their generation, and the evidence their consolidation provides about that originating body, as essential to preserving and maintaining context."[191] Jefferson Bailey and others have clearly articulated that these traditional underpinnings do not make sense with complex born digital materials.

With text mining and database availability, we can now readily search across folders and boxes, circumventing the formerly laborious "needle-in-the-haystack" search of researchers. It's now possible to not only search across whole collection, but to search across multiple collections. Non-linear retrieval supplants the narrative backbone that originated *respect des fonds.* An interesting more recent development that may challenge our traditional notions of collection-based access was recently cited this past fall as resultant of digital imaging. In September 2013, Stanford University Libraries allowed researcher to use FTK, the digital forensics software, in their lab to search across 200 removable media and 2 hard drives.[192] Cross-drive analysis also has implications for the relevance of a traditional finding aid for these types of materials. Complex materials that fall outside of text-searching are even less suitable for traditional finding

[191] Bailey, "Disrespect des Fonds."
[192] Lee, "From Bitstreams to Heritage," 31.

aids. The Timothy Leary and Salman Rushdie papers, for example, both required database access points to make sense of searching across large digital collections.

Item-level Units

Intellectual arrangement and description requires organization through series, subseries, and even item levels. With complex, inter-related, interactive, and networked media, these arrangements are less functional. It is much less helpful to think about item-level as digital files. A digital file is not a piece of paper. In the Timothy Leary papers, for example, individual digital files could be totally meaningless and inoperable when separated from other files on the disk. While Emory University's MARLB has been describing each floppy disk as a separate item for their Alice Walker collection, New York Public Library combined related whole disk images together into single access points. Several disk images may have been combined, for example, for "Disk Image 10," with each disk image containing 35-50 files each. Similar to a database or network in structure, alone each file is meaningless, indecipherable, and unimportant.

The unit question is not simply one of arrangement and description. In my discussions with Erika Farr, this is a procedural issue as well. Archives need to be able to standardize processing procedures through finding a way to talk about digital collections comparatively. One major comparison for measuring progress is by the unit level. With other materials, collections are described with units such as item, boxes, or feet, but born-digital items have no clear unit measurement except size, such as terabyte. Complex born digital collections must, as in many other problematic areas, rely on using multiple descriptions that include removable media type and number, aggregate file size, and number and types of files.

Access Copies

The process of disk imaging has implications for understanding originality as well. If an access copy and preservation copy are, on a bit-for-bit level, indistinguishable, what is an access copy? In an era in which bit-by-bit replication is easy, some collections may rethink what it means to give a researcher an access copy. In the right collection, without the worry of major restrictions, archives may think of giving out disk images to researchers for more collaborative off-site work. If the material, in analog form, could be photocopied, what biases are holding archives back from beginning to make electronic media available? Making available access copies of audio visual material, for example, is not uncommon, it remains to be seen if archives will start making bit-exact disk images also available.[193]

Opportunities for Collaboration

In addition to multi-pronged access methods and types, researcher access may also, by necessity, mean new collaborative opportunities. Born digital materials will give researchers new access to the code, the software, and timestamps preserved through digital forensics. They may require different skills and backgrounds of future researchers as well as archivists. Not only will a researcher need a content-specific background, but quite possibly he/she may need a legacy software background to best make sense of materials. Tal Nadan, the Reference Archivist at New York Public Library, envisioned a future Timothy Leary researcher accessing the material and collaborating with NYPL and Futique to not only view the games, but to build and complete Leary's unfinished games. Similarly, scholar Matthew Kirschenbaum noted in conversation with

[193] Kirschenbaum, Born-Digital Forensics seminar class discussion.

me that my own research on Timothy Leary should involve dynamic content modeling to build a visualizations of the textual content. These skill-sets are outside the confines of most humanities graduate programs as well as most archival programs. Additionally, projected collaborations between archives and researchers will depend on a much closer working relationship in order to realize their potential. By necessity, researchers as well as archivists will need to experiment as best practices are codified in this nascent area.

Running parallel to this is the emergence of digital archivist positions as institutions attempt to address the born digital materials. Like digital forensics, the figure of the digital archivist requires well-considered integration within major institutional workflows. The archival context with complex materials has shifted further and further outside traditional boundaries to include collaborating with donors, estates, artists, and researchers, as well as computer and software experts, software engineers, and digital humanities scholars. As the previous pages attest, digital archiving, forensics, and preservation require considerable collaborative work for success.

This leads to my final point: digital forensics applications in archival workflows is, if anything, a story of collaboration across not only the silos within institutions, but across major disciplinary fault lines. This was seen in the case of the BoDAR team created at Emory University in which different specialized skill-sets across multiple departments needed to work together. While most institutions will not be able to replicate such cross-disciplinary work, archives would do well to instigate partnerships with digital humanities centers to foster research-driven collaborations. Although traditionally archives have not partnered with researchers to actively work as co-collaborators, digital humanities scholars may be natural partners as users and collaborators to processing and making meaning of these little-accessed collections. Stanford University has been able to generate excitement through its pioneering work in processing and

creating visualizations from the large amount of complex born digital collections. Their customized tool will allow dynamic visualizations of networks and word usages. In my attempts to access materials thus far, I have not ever seen institutions reaching out to allow for such dynamic usage of their material.

In this emerging area in cultural archives it is still unclear if researchers are aware and accessing born digital material. One (predictive question) with important implications would be: how have recent publicized born digital collections been accessed and used? Has this had any effect on research trends? How are these materials interpreted and used in the burgeoning area of Digital Humanities? These questions are important to understand how better to provide access to complex material, but also to better advocate for their preservation and for robust workflows and scalar methods to be implemented with institutional-wide support. It still remains to be seen how new access methods to born digital work will affect research and scholarship in the future. As of now, despite the major publications on the born digital Rushdie papers, no such research making use of these collections has been published.

Through this analysis, we should be better understand the intellectual, philosophical, ethical, as well as technological challenges and implications to digital forensics. Digital forensics offers totally new opportunities, but it also makes it possible to side-step long-held archival principles. The future of access to born digital archives bodes a different understanding of intellectual objects and groupings. It remains to be seen how archives will makes sense of these challenges. With the acceptance of digital forensics into the nomenclature, however, it is insufficient for major institutions to accept born digital materials without any disk imaging procedures in place.[194] Rushdie and Leary are only the very beginning. Developments in

194 Jeonghyun Kim, Edward Warga, and William Moen. "Digital Curation in the Academic Job Market." ASIST 2012 October 28-31, 2012. Baltimore, MD, https://www.asis.org/asist2012/proceedings/Submissions/283.pdf.

frameworks, tools, and workflows, institutions with robust digital preservation programs should be able to take steps to image legacy media before it is too late.

Appendix

Emulation versus Migration

Erik Oltmans and Nanda Kol, "A Comparison Between Migration and Emulation in Terms of Costs," RLG DigiNews, vol. 9, no. 2, http://www.rlg.org/en/page.php?Page_ID=20571#article0 (accessed on 1 May 2010);

Stewart Granger, "Emulation as a Digital Preservation Strategy," DLib Magazine, vol. 6, no. 10, http://www.dlib.org.proxy.library.emory.edu/dlib/october00/granger/10granger.html (accessed on 1 May 2010); Koninklijke Bibliotheek, "Emulate," http://www.kb.nl/hrd/dd/dd_projecten/projecten_emulatie-en.html (accessed on 15 May 2010).

Gregory Miura, "Emulation Expert Meeting," International Preservation News 40 (December 2006), pp. 39–40, http://archive.ifla.org/VI/4/ipn.html (accessed on 1 May 2010).

Jeffery Rothenberg, "Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation" (1999), http://www.clir.org/pubs/abstract/pub77.html (accessed on 11 July 2011).

Instructional

Erin Kenneally. 2001 "Gatekeeping out of the box: open source software."Viginia Journal of Law and Technology. http://www.vjolt.net/vol6/issue3/v6i3-a13-Kenneally.html

Ewa Huebner, Open Source Software for Digital Forensics," (Springer 2010) doi 10.1007/978-1-4419-5803-7. Accessed November 17, 2013. http://ezproxy.library.nyu.edu:2178/book/10.1007%2F978-1-4419-5803-7.

Julianna Barrera-Gomez and Ricky Erway, "Walk This Way: Detailed Steps for Transferring Born digital Content from Media You Can Read In-House. Dublin, Ohio: OCLC Research. http://www.oclc.org/content/dam/research/publications/library/2013/2013-02.pdf

Julianna Barrera-Gomez,"You've Got to Walk Before You Can Run: First Steps for Managing Born digital Content Received on Physical Media."

Porter Olsen, "Digital Curation Workstation," MITH. online

The Enigma Email. Digital forensics personal archives http://lib.stanford.edu/born digital-program-stanford/enigma-email
We then collaborated with Elijah Meeks, a digital humanities specialist at Stanford University Libraries, who used Robert Creeley's email headers in Gephi to illustrate the organic network visualizations of the poet 50,000 emails. However, while it illustrated the frequency and relationships of senders and recipients, the results lacked expression of these complex relationships in combination with topics/subjects.

Digital Forensics

Chris Prom. 2012. “Characterizing Files” Practical E-Records: Software and Tools for Archivists (blog)

Dan Farmer and Wietse Venema, Forensics Discovery (New York: Knopf, 2007), 59-51.

Forensics Wiki, accessed November 16, 2013.
http://www.forensicswiki.org/wiki/Main_Page

Garfinkel, Simson

Gareth Knight, “The Forensics Curator: Digital Forensics as a Solution to Addressing the Curatorial Challenges Posed by Personal Digital Archives,” International Journal of Digital Curation, 2012, 7:2, 40-63. November 16, 2013.
http://www.ijdc.net/index.php/ijdc/article/view/218

Jeremy L. John, 2012. Digital Forensics and Preservation. DPC Technology Watch Report 12-03. Great Britain. Digital Preservation Coalition.
http://www.dpconline.org/advice/technology-watch-reports

Paul Nicholls *Multimedia: A Management Perspective,* Antone F. Alber, (ed.) (elmont CA: Wadsworth Publishing Company, 1996)242.

Matthew Kirschenbaum, R Ovenden, Gabriela Redwine(2010).Digital Forensics and Born digital Content in Cultural Heritage Collections. Council on Library and Information Sources:http://www.clir.org/pubs/abstract/reports/pub149

Matthew Kirschenbaum. The Signal. interview
http://blogs.loc.gov/digitalpreservation/2013/08/whats-a-nice-english-professor-like-you-doing-in-a-place-like-this-an-interview-matthew-kirschenbaum/

Matthew Kirschenbaum. (2009). Where Computer Science and Cultural Studies Collide.Chronicle of Higher Education, 55(20), B11.

Matthew Kirschenbaum. New Media and the Forensic Imagination (MIT Press, 2008).

Matthew Kirchenbaum. “Mechanisms: New Media adn ntheh Forensic Imagination”

Seamus Ross and Ann Gow (1999), Digital Archaeology: Rescuing Neglected and Damaged Data: A JISC/NPO Study within the Electronic Libraries (eLib) Programme on the Preservation of Electronic Materials, (Glasgow, Humanities Advance Technology and Information Institute:

1999). Online. Accessed November 17, 2013.
http://www.ukoln.ac.uk/services/elib/papers/supporting/pdf/p2.pdf

Digital Preservation and Archiving

James Currall and Michael Moss, "We are Archivists, but are we OK?," Records Management Journal 18.1 (2008): 69-91, accessed November 11, 2013.
http://ezproxy.library.nyu.edu:2128/docview/223829420/fulltext?source=fedsrch&accountid=12768
Shiralee Saul, who sees the internet as "the Wun'derkammer to end all Wunderkammer"

Kenneth Thibodeau, "Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years," in The State of Digital Preservation; Garrett and Waters, Preserving Digital Information

Margaret Hedstom. "Digital Preservation: The Bomb for Digital Library a Computers http://www.uky.edu/~kiernan/DL/hedstrom.html

Naomi L. Nelson, Managing born digital special collections and archival materials (city: Association of Research Libraries, 2012).

Richard Pearce-Moses, "The Perfect and the Possible: Becoming a Digital Archivist," paper presented at the Conference of Inter-Mountain Archivists, 12 May 2006, available at http://www.lib.az.us/about/annualreports

Susan Davis. "Electronic Records Planning in 'Collecting' Repositories," *American Archivist. 2007*

http://blogs.loc.gov/digitalpreservation/category/tools-and-infrastructure/

http://dpoutreach.net/

DH as collaborative, focused on building useable tools, based on the internet.

Personal Digital Archive Case Studies

Andy, "Pressed into the Service of Cinema: Issues in Preserving the Software of Hollis Frampton and the Digital Arts Lab," The Moving Image 12:1 (2012): 18-43. Online http://ezproxy.library.nyu.edu:2255/ehost/pdfviewer/pdfviewer?sid=e7b214d2-3e25-4408-94b9-4e091bcca469%40sessionmgr10&vid=2&hid=10

Greg Miller. "Turn On, Boot-Up, and Jack in With Timothy Leary's Long-Lost Videogames." *Wired*. October 1, 2013. online. http://www.wired.com/wiredscience/2013/10/timothy-leary-video-games/

Laura Carroll, Erika Farr, Peter Hornsby, Ben Ranker. "A Comprehensive Approach to Born - Digital Archives." *Archivaria* vol 72 (2011): 61-92. Accessed November 30, 2013. http://pid.emory.edu/ark:/25593/cksgv .

Martin G. Gengenbach,. "The Way We Do It Here" Mapping Digital Forensics Workflows in Collecting Institutions. A Master's Paper for the M.S. in L.S degree. August, 2012. 131 pages. Advisor: Christopher A. Lee. http://digitalcurationexchange.org/system/files/gengenbach-forensic-workflows-2012.pdf

Mary J. Loftus. "The Author's Desktop" Emory Magazine. Winter 2010. online http://www.emory.edu/EMORY_MAGAZINE/2010/winter/authors.html

Patricia Cohen, "Fending Off Digital Decay, Bit by Bit." New York Times. March 16, 2010. online. http://www.nytimes.com/2010/03/16/books/16archive.html?pagewanted=all

Patricia Cohen. New York Times. June 15, 2011, online. http://www.nytimes.com/2011/06/16/books/new-york-public-library-buys-timothy-learys-papers.html?pagewanted=all.

"Playing Timothy Leary's Lost Games." Kotaku. November 5, 2013. You Tube. http://www.engadget.com/2013/09/30/timothy-leary-video-game-archive-drugs-are-bad-mmkay/

Trevor Owens. "Born Digital Archival Materials at NYPL: An Interview with Donald Mennerich," *The Signal: Digital Preservation*, April 22, 2013. Library of Congress. Accessed November 10, 2013.

Susan Thomas, Renhart Gittens, Janette Martin, and Fran Baker, Workbook on Digital Private Papers, 2005–2007, Paradigm Project, available at http://www.paradigm.ac.uk/ workbook/index.html (accessed 12 February 2011).

Douglas Elford, Nicholas Del Pozo, Snezana Mihajlovic, David Pearson, Gerard Clifton, and Colin Webb, "Media Matters: Developing Processes for Preserving Digital Objects on Physical Carriers at the National Library of Australia," World Library and Information Congress: 74th IFLA General Conference and Council, 10–14 August 2008, available at http://www.ifla.org/IV/ifla74/papers/084-Webb-en.pdf (accessed on 12 February 2011);

Forstrom; Catherine Stollar Peters, "When Not All Papers are Paper: A Case Study in Digital Archivy," Provenance, vol. XXIV (Atlanta, 2006), available at https://ford.ischool. utexas.edu/bitstream/2081/2226/1/023-035.pdf (accessed on 12 February 2011);

Catherine Stollar and Thomas Kiehne, "Guarding the Guards: Archiving the Electronic Records of Hypertext Author Michael Joyce," New Skills for the Digital Era, Case Study 4, available at http://www.archivists.org/publications/proceedings/NewSkillsForADigitalEra.pdf (accessed on 12 February 2011);

Chris Hilton and Dave Thompson, "Collecting Born Digital Archives

at the Welcome Library," Ariadne 50 (30 January 2007), available at http://www.ariadne.

Appendix B: The Digital Archives of Salman Rushdie: Overview

Overview

Welcome to the Salman Rushdie digital archive. On this workstation, you will find selected digital files from Rushdie's Macintosh Performa 5400, one of several computers and other related devices that form the born digital component of the Salman Rushdie papers in Emory University's Manuscript, Archives, and Rare Book Library (MARBL).

The majority of the digital files date from 1992-2002, and consists of notes and drafts of Rushdie's writings and selected correspondence. Of particular interest is a small cache of email correspondence, representing Rushdie's first foray into this emerging form of communication in the late 1990s. Writings include drafts of Rushdie's fiction, such as *East, West* (1994), *The Moor's Last Sigh* (1995), and *The Ground Beneath Her Feet* (1999). Nonfiction writings include notes and drafts for *Step Across this Line*, Rushdie's collection of essays and criticism, published in 2002. Other writings include drafts of the *Midnight's Children* and *The Courter* play scripts, as well as drafts of letters to the editor, newspaper columns, poems, and speeches. The Performa 5400 contains a backup of an earlier computer, which Rushdie entitled "OLD MAC," and a laptop, the "Powerbook," which Rushdie likely used in tandem with the Performa 5400.

Finding Aid

Note that additional drafts of Rushdie's writings may be found in the non-digital portion of the Rushdie papers. Consult the Salman Rushdie papers finding aid for an overview of the entire collection and a detailed box and folder listing for the paper portion of the collection. The born digital materials are also briefly described to provide a complete picture of the nature and extent of the entire collection.

Searching and Using the Born Digital Files

- **Emulated Environment**

  Start here to experience the look and feel of Rushdie's original computer environment. Emulation employs current technologies to imitate older systems and applications. In this environment, you will be able to view Rushdie's exact directory structure and open each file in the application in which it was created, such as MacWrite Pro or ClarisWorks.

- **Searchable Database**

  Use the searchable database to find a specific text or to browse the Performa 5400 by directory or type of work, such as fiction, nonfiction, or correspondence. This database enables you to conduct simple or advanced full text searches of individual files and displays the majority of files in PDF format; therefore some original formatting and context may be lost.

Special restrictions

Material available on this workstation is for research use only. Researchers are not permitted to copy or download any files or applications from the emulated environment or the searchable database.

US Copyright Law (Title 17, U.S. Code) provides protection for published and unpublished materials. To publish or distribute copyrighted materials, the copyright owner must give permission, especially for previously unpublished materials. It is the researcher's responsibility to secure that permission.

---

http://marbl.library.emory.edu/documents/digital-archives/rushdie-help-sheet.pdf

Appendix C:

Neurostar				
	photos for neurmancer roles			
Roles	**Burroughs**	**Gibson**	**Pynchon**	**Leary**
Case	gay punk	doper-punk	horney guy	Young scientist
david bowie	david bowie	david byrne	david keith	peter'otoole
	kim?	Case	Blothrop	leary
Molly	sara doublas	grace jones	sus surandon	barbara
	---	Molly	Katje	barbara
Linda	karl lagerfield	madonna (circled)	maryln monroe	tina chow
(whore w/heart of gold)	---	Linda	Witch	maia
Riviera	islam jihad	Turings	Pisces	DEA
Cop	Cord Meyer	Yves St. Laurant	Edwin Meese	Gordon Liddy
	---	Riviera	majro mary	Gordon Liddy
Deane	Burroughs			
	Burroughs			
Finn				
Voodoo				
3-jane	raquel welch			
3 janDad	chris isherwood			
Musician	Lou reed	rolling stones	beatles	talking heads
	Dylan			
good team				
Co mix				

master matrix of a mind movie				
Track	A	B	C	D
	Gibson	Pynchon		
Act 1 scene 1				
Location	chat bar	Brothel	Pablos	west. Bar
Case	Case	Slothrop	dedalus	Kim
Star	Jagger	G toole	byrne	Bowi
Them				

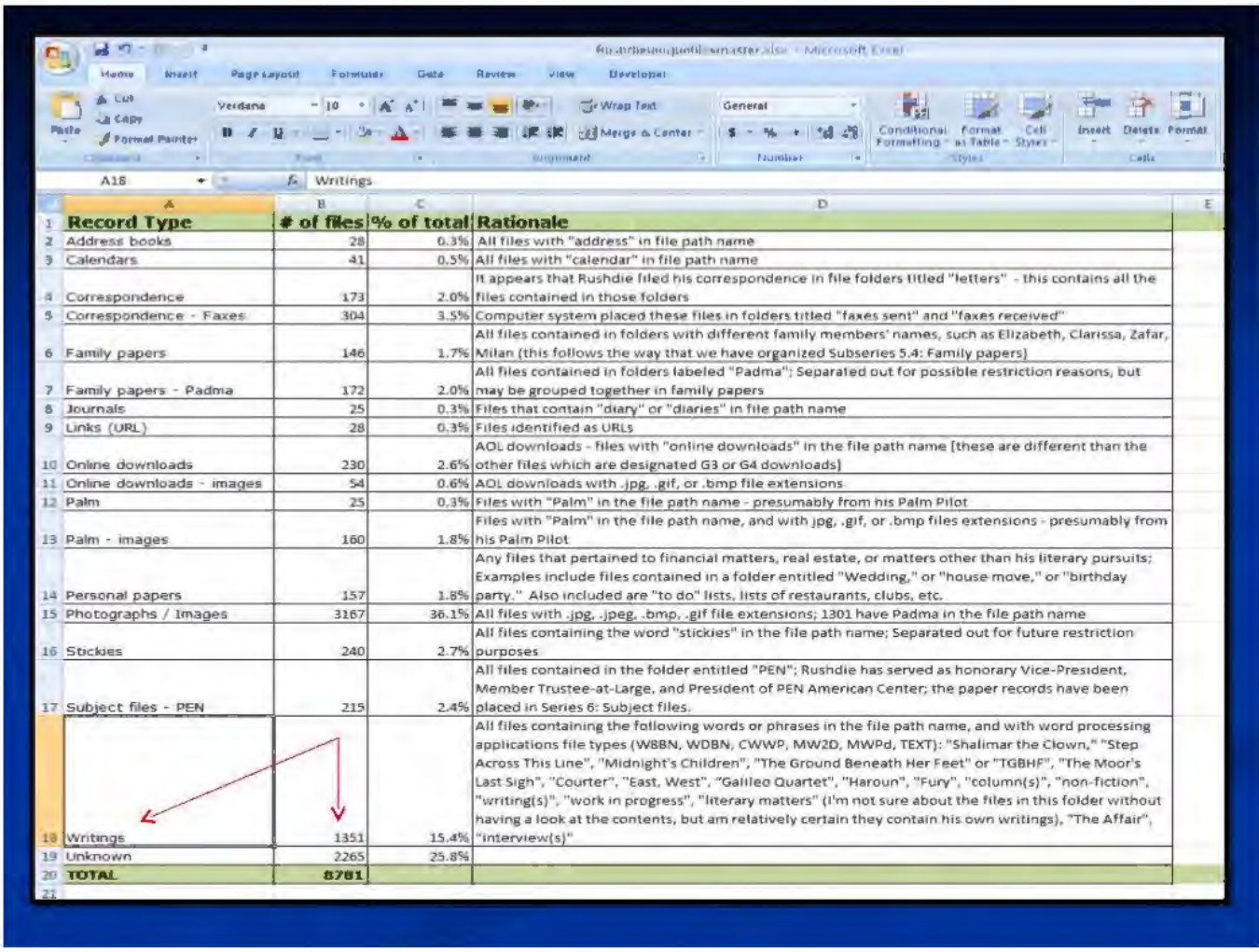
Act 1 scene 2				
Location	chat bar	Brothel	pablo's	west bar
Linda	Linda	spanish M	m. bloom	---
star theme	Madonna	Marilyn M	tina chow	----

Act 1 scene 3				
Location	coffin hot.	Nice hotel	hotel	Bunker
Molly	molly'd. hanna	Katje	bar	Bar
Star		ang. Houston	bar	Bar
Theme				

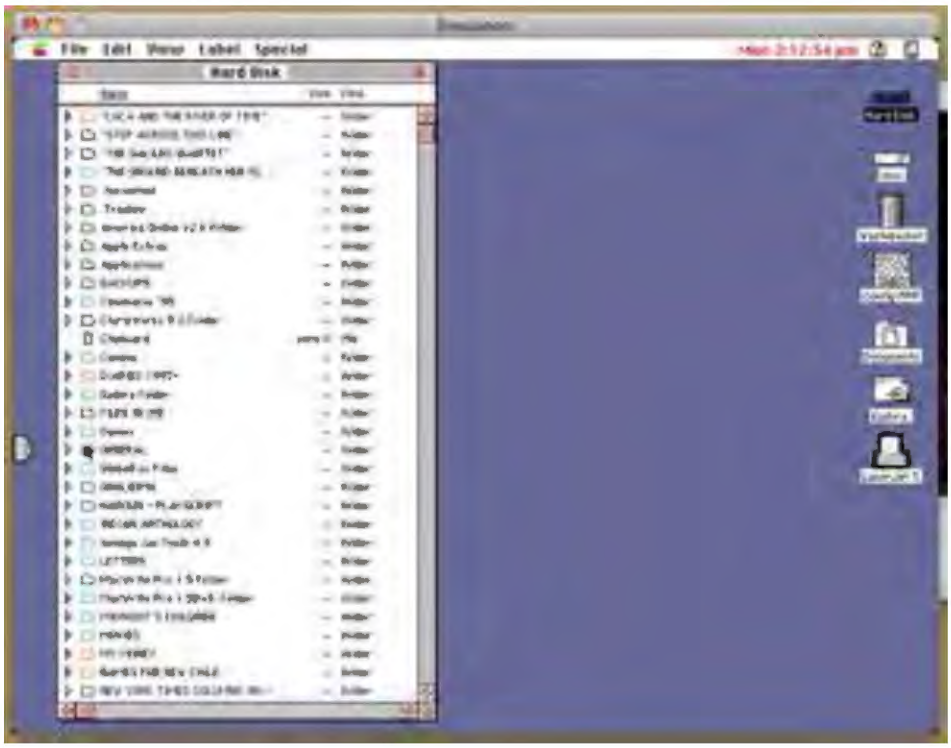
Act 1 scene 4				
Location	finn-shop	Berlin	psych lab	gun shop
Finn	finn-shop	Bodine	Huxley	w.b. Dad
Star	Isherwood	Gaddis	Huxley	Phil Dick
Theme				


NEURO
MANCER
MINDMOVIE

Neuromancer screenshots, New York Public Library



Record Type	# of files	% of total	Rationale
Address books	28	0.3%	All files with "address" in file path name
Calendars	41	0.5%	All files with "calendar" in file path name
Correspondence	173	2.0%	It appears that Rushdie filed his correspondence in file folders titled "letters" - this contains all the files contained in those folders
Correspondence - Faxes	304	3.5%	Computer system placed these files in folders titled "faxes sent" and "faxes received"
Family papers	146	1.7%	All files contained in folders with different family members' names, such as Elizabeth, Clarissa, Zafar, Milan (this follows the way that we have organized Subseries 5.4: Family papers)
Family papers - Padma	172	2.0%	All files contained in folders labeled "Padma"; Separated out for possible restriction reasons, but may be grouped together in family papers
Journals	25	0.3%	Files that contain "diary" or "diaries" in file path name
Links (URL)	28	0.3%	Files identified as URLs
Online downloads	230	2.6%	AOL downloads - files with "online downloads" in the file path name [these are different than the other files which are designated G3 or G4 downloads]
Online downloads - images	54	0.6%	AOL downloads with .jpg, .gif, or .bmp file extensions
Palm	25	0.3%	Files with "Palm" in the file path name - presumably from his Palm Pilot
Palm - images	160	1.8%	Files with "Palm" in the file path name, and with jpg, .gif, or .bmp files extensions - presumably from his Palm Pilot
Personal papers	157	1.8%	Any files that pertained to financial matters, real estate, or matters other than his literary pursuits; Examples include files contained in a folder entitled "Wedding," or "house move," or "birthday party." Also included are "to do" lists, lists of restaurants, clubs, etc.
Photographs / Images	3167	36.1%	All files with .jpg, .jpeg, .bmp, .gif file extensions; 1301 have Padma in the file path name
Stickies	240	2.7%	All files containing the word "stickies" in the file path name; Separated out for future restriction purposes
Subject files - PEN	215	2.4%	All files contained in the folder entitled "PEN"; Rushdie has served as honorary Vice-President, Member Trustee-at-Large, and President of PEN American Center; the paper records have been placed in Series 6: Subject files.
Writings	1351	15.4%	All files containing the following words or phrases in the file path name, and with word processing applications file types (WBBN, WDBN, CWWP, MW2D, MWPd, TEXT): "Shalimar the Clown," "Step Across This Line", "Midnight's Children", "The Ground Beneath Her Feet" or "TGBHF", "The Moor's Last Sigh", "Courter", "East, West", "Galileo Quartet", "Haroun", "Fury", "column(s)", "non-fiction", "writing(s)", "work in progress", "literary matters" (I'm not sure about the files in this folder without having a look at the contents, but am relatively certain they contain his own writings), "The Affair", "interview(s)"
Unknown	2265	25.8%	
TOTAL	8781		



https://scholarblogs.emory.edu/woodruff/fyi/the-digital-archives-program-settles-in-to-marbl