

Elective Course Paper

Special Collections Librarianship and The Future of Born-Digital Materials

Written for IS-439: *Special Collections* with Prof. Anna Chen, Winter 2019

Assigned to “reflect on some aspect of special collections librarianship,” in this short paper, I built upon my research into emulation technologies to imagine a new, “post-codex” future for special collections. While most of our course readings dealt with rare books, unique manuscripts and paper artifacts, I used this opportunity to investigate my overlapping interests in digital preservation and accessibility of special collections materials. Highlighting the materiality of born-digital materials, this paper pairs well with my portfolio issue paper to express my growing interest in the preservation of complete, “authentic” representations of obsolete media. In addition to a brief discussion of emulation, I used a “media archeology” lens to highlight the importance of the physicality of computing hardware and hoped to show that archival representation of born-digital materials remains an important and under-explored topic within special collections. A lengthier treatment of this subject would benefit from more in-depth case-studies of specific technologies and digital objects and the costs and feasibility of incorporating “retro-computing” into reading rooms.

Special Collections Librarianship and The Future of Born-Digital Materials

By considering current tools in digital preservation and examining emergent trends in access to born-digital materials, in this paper, I will identify the shifting priorities and core knowledge requirements of special collections librarianship. This paper suggests that special collections libraries at academic research institutions should devote more resources to the collection, preservation, and documentation of obsolete born-digital media and related historic computing equipment.

Over the last two decades, as born-digital materials have begun to flood into archives and special collections, most institutions have primarily focused their resources on accurately ingesting and preserving these records, while paying little attention to the means by which they would eventually be accessed.¹ By migrating material off of obsolete carriers like floppy disks and CD-ROMs, digital preservationists have saved otherwise at-risk cultural artifacts and staved off the inevitable physical deterioration of these resources. The increasing availability and reliability of digital forensics tools such as checksums, write blockers,² and BitCurator³ has made this task more reliable and standardized, but a major issue still looms: many of these preserved files are no longer readable by contemporary computers. Moreover, due to the rapid evolution and increasingly proprietary nature of file formats, the files that are still readable require highly specialized or obsolete software to properly run.⁴ While there are many programs that offer backwards compatibility with older file formats, recent research has shown that “[w]hen files are

¹ Julia Kim, “Researcher Access to Born-Digital Collections: An Exploratory Study,” *Journal of Contemporary Archival Studies* 5, no. 1 (May 25, 2018), <https://elischolar.library.yale.edu/jcas/vol5/iss1/7>.

² Matthew G. Kirschenbaum et al., *Digital Forensics and Born-Digital Content in Cultural Heritage Collections*, CLIR Publication, no. 149 (Washington, D.C: Council on Library and Information Resources, 2010).

³ “<https://bitcurator.net/>,” accessed March 16, 2019, <https://bitcurator.net/>.

⁴ Jessica Meyerson et al., “The Software Preservation Network (SPN): A Community Effort to Ensure Long Term Access to Digital Cultural Heritage,” *D-Lib Magazine* 23, no. 5/6 (May 2017), <https://doi.org/10.1045/may2017-meyerson>.

rendered in environments that differ from the original then they will often present altered information to the user.”⁵ For institutions dedicated to authenticity and trust, such as libraries and archives, these alterations must be taken seriously. So, the question remains, how will libraries and archives provide researchers with access to their growing collections of born-digital materials?

In his 2014 keynote address at “Acknowledging the Past, Forging the Future,” Stephen Ennis recognizes the ever-increasing importance of born-digital materials and notes that “[w]hat is most needed now is not tools, I think, but policies that will guide the access and use of this growing resource.”⁶ While Ennis is certainly correct to draw attention to the problem of researcher access, the division he draws between tools and policies is not quite as clear as he indicates. The policies for providing access to born-digital collections must be guided by a clearer understanding of the tools that are necessary to properly do this task. And the nature of those tools is a matter that is still very much undecided. Special collections librarians must begin to more seriously reckon with the question: what does access to born-digital material actually look like?

One of the most promising and rapidly evolving ways to provide this access is through the use of emulation. A full technical explanation of emulation is beyond the scope of this paper, but essentially it is a means by which a “virtual machine” is implemented on a host computer. This implementation can then run “unmodified software binaries” designed for the system

⁵ Euan Cochrane, “Rendering Matters-Report on the Results of Research into Digital Object Rendering,” *Archives New Zealand*, 2012.

⁶ case, *Objects of Study: Special Collections in an Age of Digital Scholarship: Stephen Ennis*, accessed March 20, 2019, <https://www.youtube.com/watch?v=DofAdsej7Yg&t=0s&index=7&list=PLBELrG1nZ2U5jXND2u48h4RScCZwn gM5g>.

running on that virtual machine.⁷ In other words, by virtually imitating a specific configuration of hardware and “machine code,” emulation allows for software code to be run on computers on which it could normally never run. Emulation is not a new technology, and it has been used for preservation purposes dating back to at least 1995, but due to its high cost and required level of technical expertise there have until very recently been few examples of its use in special collections.⁸

The earliest successful and most highly publicized use of emulation by a research library in the United States was Emory University’s processing of the Salman Rushdie papers in 2007.⁹ ¹⁰ Through negotiations with Rushdie, in late 2006, Emory University’s Manuscript, Archives, and Rare Book Library (MARBL) acquired, along with his traditional paper archives, “a nearly complete record of Rushdie’s digital life, consisting of four computers...one hard drive...and several disks.”¹¹ Rather than simply migrating the digital materials off of the computers to store as discrete files, MARBL created a disk image of Rushdie’s earliest computer hard drive, a Performa 5400, and then created an emulated version of this exact machine on a researcher workstation in their reading room. After processing, removing, and reloading some of the original files, this allowed researchers to “launch an exact replica of Rushdie’s Performa 5400 with all of its authentic and at times unstable, mid-1990s Mac attributes.”¹² By allowing researchers access to the contextual environment of Rushdie’s born-digital records, MARBL drew attention to the historical

⁷ David S.H. Rosenthal, *Emulation & Virtualization as Preservation Strategies*, a report commissioned by The Andrew W. Mellon Foundation, New York, October 2015, <https://mellon.org/Rosenthal-Emulation-2015>

⁸ Rosenthal, “Emulation & Virtualization as Preservation Strategies.”

⁹ Salman Rushdie papers, Stuart A. Rose Manuscript, Archives, and Rare Book Library, Emory University.

¹⁰ Dan Rockmore, “The Digital Life of Salman Rushdie,” July 29, 2014, <https://www.newyorker.com/tech/annals-of-technology/digital-life-salman-rushdie>.

¹¹ Laura Carroll et al., “A Comprehensive Approach to Born-Digital Archives,” *Archivaria* 72 (2011): 61–92, 64.

¹² Carroll et al., 84

relevance and materiality of software, and pointed to a new future for born-digital personal paper collections. In a 2010 article, Carroll et al. describe the reasoning behind their use of emulation:

Ultimately, we decided that the context and medium of twentieth and twenty-first century archives are of equal importance as those of pre- and post-Gutenberg collections. Scholarly interest in incunabula, early publishing practices, bindings, paper, manuscript hands, marginalia and back matter surely will be mirrored in scholarly research into literary and creative production in the late twentieth century and on. Identifying, categorizing, preserving, and providing access to the materiality of born-digital personal archives can be of equal importance as attending to the content, depending upon the extent, medium and state of a given collection.¹³

In many ways, Carroll et al.'s predictions have already come to fruition, as there has been a recent proliferation of scholarship on the historical and cultural importance of software¹⁴ and the emergent field of media archaeology has placed new attention on the materiality of the digital.¹⁵ By providing researchers with “the full digital paratext – the native environment and context,” Carroll et al. hoped to “enabl[e] research not only into the content but also the technological medium itself and how it might impact literary production.”¹⁶ While it is likely still too early to tell the full impact of this newly available resource, the immense research value of emulation is clear. Moreover, building off the successes of MARBL, several more recent projects have also effectively used emulation to provide access to born-digital special collections materials, most notably New York Public Library's processing of video games on floppy disks in the Timothy Leary Papers and New York University's treatment of 1990s era Photoshop files from the Jeremy Blake Papers.¹⁷

¹³ Carroll et al., 79

¹⁴ For example see the “Software Studies” series from The MIT Press, accessed March 15, 2019, <https://mitpress.mit.edu/books/series/software-studies>.

¹⁵ For example, see https://monoskop.org/Media_archaeology for a list of recent texts and scholarly courses

¹⁶ Carroll et al., “A Comprehensive Approach to Born-Digital Archives,” 82

¹⁷ Dianne Dietrich et al., “How to Party Like It's 1999: Emulation for Everyone,” *The Code4Lib Journal*, no. 32 (April 25, 2016), <http://journal.code4lib.org/articles/11386>.

Despite these successes, emulation alone cannot solve the problems of born-digital access. MARBL's innovative use of emulation to provide access to the born-digital materials in the Salman Rushdie Papers is still lacking in one major way: the physical hardware of the Performa 5400 is removed from the researcher's experience. In order to preserve the original computer, researchers must access the files through the hardware of the workstation computer in the reading room. This points to one of the major drawbacks of using emulation as an access tool. As director of the Media Archeology Lab (MAL) Lori Anderson explains in reference to her work on early electronic literature, "Emulation will never be able to replicate the entire physical, sensory, tactile experience of working on the original machine. The clackety-clack of the keyboard, the act of taking the 5.25" floppy out of its sleeve, sliding it into the drive, hearing the whir and beep of the machine, the ability to open up the hood and insert an expansion card is integral to the reading/writing experience."¹⁸ Despite MARBL's commitment to the "full digital paratext," the emulation is still lacking in its representation of the physical, material reality of Rushdie's computing environment. To more accurately and truly represent this full context, the library would need to have a version of the old Performa 5400 for access use. Obviously, with the vast and ever-expanding variations of computers that have been produced throughout the last 50 years, no institution could be expected to own a comprehensive or even significant portion of this set. Nevertheless, there are a growing number of institutions, both amateur and professional, that are committed to maintaining the material history of computing, and special collections libraries would be wise to look to these groups for

¹⁸ Trevor Owens, "Media Archaeology and Digital Stewardship: An Interview with Lori Emerson | The Signal," webpage, October 11, 2012, <https://doi.org//blogs.loc.gov/thesignal/2012/10/media-archaeology-and-digital-stewardship-an-interview-with-lori-emerson/>.

collaboration and guidance.¹⁹ For example, recent work by John Durno at the University of Victoria Libraries has shown the immense value of partnering with amateur enthusiasts to access and recover data stored on floppy disks from the 1980's.²⁰

Driven primarily by private collectors and amateur enthusiasts, “retrocomputing” is rapidly emerging as a valuable and promising source of new research and scholarship. In his 2013 essay “Ancient Evenings: Retrocomputing in the Digital Humanities” Mathew Kirschenbaum demonstrates the value of maintaining obsolete computing environments within the growing academic discipline of digital humanities. He defines retrocomputing as “a set of hands-on practices devoted to preserving, engaging, and extending the historical legacy of outdated and outmoded computer systems for purposes of documentation and recovery, education, experimentation, critical and artistic expression, and sheer satisfaction.”²¹ While Kirschenbaum never draws a direct comparison to the history of rare book collecting, the parallels are clear. And just as rare book collectors, enthusiasts and hobbyists played an essential role in the development of many special collections, so too should the retrocomputing community. Noting this potentially important role for the retrocomputing community, Yuri Takhteyev and Quinn DuPont point out that “the collection and preservation efforts of retrocomputing enthusiasts ... fill an important gap in the work of traditional memory institutions.”²² This important gap cannot be filled by any single institution alone, nor will it ever

¹⁹ Patricia Galloway, “Retrocomputing, Archival Research, and Digital Heritage Preservation: A Computer Museum and ISchool Collaboration,” *Library Trends* 59, no. 4 (2011): 623–36, <https://doi.org/10.1353/lib.2011.0014>.

²⁰ John Durno and University of Victoria Libraries, “Digital Archaeology and/or Forensics: Working with Floppy Disks from the 1980s,” *The Code4Lib Journal*, no. 34 (October 25, 2016), <http://journal.code4lib.org/articles/11986>.

²¹ Matthew G. Kirschenbaum, “Ancient Evenings,” in *A New Companion to Digital Humanities* (John Wiley & Sons, Ltd, 2015), 185–98, <https://doi.org/10.1002/9781118680605.ch13>.

²² Yuri Takhteyev and Quinn DuPont, “Retrocomputing as Preservation and Remix,” *Library Hi Tech* 31, no. 2 (June 7, 2013): 355–70, <https://doi.org/10.1108/07378831311329103>.

be fully filled (there will always be gaps and absences), but special collections libraries can and should play an important part in that attempt.

Because special collections and rare books librarians' have such a long-held commitment to preserving and documenting the material and physical qualities of textual objects, they should be well suited to adapt their skills to considering new forms of text that transcend the traditional codex. Despite their new form, born-digital objects are, like books, physical things, and the ways in which we should preserve and steward these objects for future use is still very much taking form. Special collections libraries must develop strategies that center the physical materiality of digital objects. Research institutions should seek to maintain a more authentic version of the embodied, sensory way born-digital materials are used. In other words, if the value of special collections research is going to transcend the era of the printed book, librarians must begin learning and documenting the new rapidly disappearing forms of cultural production.

Bibliography

- Carroll, Laura, Erika Farr, Peter Hornsby, and Ben Ranker. "A Comprehensive Approach to Born-Digital Archives." *Archivaria* 72 (2011): 61–92.
- case. *Objects of Study: Special Collections in an Age of Digital Scholarship: Stephen Enniss*. Accessed March 20, 2019. <https://www.youtube.com/watch?v=DofAdsej7Yg&t=0s&index=7&list=PLBELrG1nZ2U5jXND2u48h4RScCZwngM5g>.
- Cochrane, Euan. "Rendering Matters-Report on the Results of Research into Digital Object Rendering." *Archives New Zealand*, 2012.
- Dietrich, Dianne, Julia Kim, Morgan McKeehan, and Alison Rhonemus. "How to Party Like It's 1999: Emulation for Everyone." *The Code4Lib Journal*, no. 32 (April 25, 2016). <http://journal.code4lib.org/articles/11386>.
- Durno, John, and University of Victoria Libraries. "Digital Archaeology and/or Forensics: Working with Floppy Disks from the 1980s." *The Code4Lib Journal*, no. 34 (October 25, 2016). <http://journal.code4lib.org/articles/11986>.
- Galloway, Patricia. "Retrocomputing, Archival Research, and Digital Heritage Preservation: A Computer Museum and ISchool Collaboration." *Library Trends* 59, no. 4 (2011): 623–36. <https://doi.org/10.1353/lib.2011.0014>.
- "<https://bitcurator.net/>." Accessed March 20, 2019. <https://bitcurator.net/>.
- Kim, Julia. "Researcher Access to Born-Digital Collections: An Exploratory Study." *Journal of Contemporary Archival Studies* 5, no. 1 (May 25, 2018). <https://elischolar.library.yale.edu/jcas/vol5/iss1/7>.
- Kirschenbaum, Matthew G. "Ancient Evenings." In *A New Companion to Digital Humanities*, 185–98. John Wiley & Sons, Ltd, 2015. <https://doi.org/10.1002/9781118680605.ch13>.
- Kirschenbaum, Matthew G., Richard Ovenden, Gabriela Redwine, and Rachel Donahue. *Digital Forensics and Born-Digital Content in Cultural Heritage Collections*. CLIR Publication, no. 149. Washington, D.C: Council on Library and Information Resources, 2010.
- Meyerson, Jessica, Zac Vowell, Wendy Hagenmaier, Aliza Leventhal, Elizabeth Russey Roke, Fernando Rios, and Tim Walsh. "The Software Preservation Network (SPN): A Community Effort to Ensure Long Term Access to Digital Cultural Heritage." *D-Lib Magazine* 23, no. 5/6 (May 2017). <https://doi.org/10.1045/may2017-meyerson>.
- Owens, Trevor. "Media Archaeology and Digital Stewardship: An Interview with Lori Emerson | The Signal." Webpage, October 11, 2012. <https://doi.org//blogs.loc.gov/thesignal/2012/10/media-archaeology-and-digital-stewardship-an-interview-with-lori-emerson/>.

Press, The MIT. "Software Studies." The MIT Press. Accessed March 20, 2019.
<https://mitpress.mit.edu/books/series/software-studies>.

Rockmore, Dan. "The Digital Life of Salman Rushdie," July 29, 2014.
<https://www.newyorker.com/tech/annals-of-technology/digital-life-salman-rushdie>.

Takhteyev, Yuri, and Quinn DuPont. "Retrocomputing as Preservation and Remix." *Library Hi Tech* 31, no. 2 (June 7, 2013): 355–70. <https://doi.org/10.1108/07378831311329103>.