

File not Found: Rarity in the Age of Digital Plenty

Doug Reside

A pre-publication draft of an article first published in the spring 2014

issue of RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage

In the first section of the submission guidelines for this esteemed journal, would-be authors are informed, “*RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage* uses a web-based, automated, submission system to track and review manuscripts. Manuscripts should be sent to the editor, [...], through the web portal[...].” The multivalent uses of the word, “manuscript” in this sentence reveal a good deal about the state of our field. This journal is dedicated to the study of manuscripts, and it is understood by most readers that the manuscripts being studied are of the “one-of-a-kind” variety (even rarer than the “rare books” that are also within the journal’s scope). Yet within the same sentence, the sense of the word changes to include the easily replicated digital objects that might be “tracked and review[ed]” using a “web-based, automated, submission system” and “sent [...] through a web portal.” There are clearly many formal differences between the digital files submitted through the RBM web portal and the holograph copy of *Frankenstein* Mary Shelley sent to her publisher, but they

share enough similarities that, in English at least, the same word is used to describe both.

Taken literally, the idea of a digital manuscript, where “manuscript” retains its etymological meaning of a handwritten document, is something of an oxymoron. Perhaps digital files that record the movement of a hand-held digital pen or stylus, or early data files encoded by hand on punch tape or cards, may qualify, but such artifacts represent only a small fraction of what are today considered digital manuscripts. Of course, manuscript divisions have long ago broadened their scope to include typescripts and other mechanically produced documents that record a writer’s creative process, and the word “manuscript” may now mean something like “unpublished drafts created by the author.”

A typescript, however, is more similar to a manuscript than either is to a digital manuscript. Whether a letter is inscribed onto paper by an engraved inked mallet or sketched by an inked stylus, the physical connection of the author’s hand to the written word is a close one in the analog world. Not so today when the strike of a key invokes a series of encodings causing pixels to glow, for a moment, in a letter-shaped pattern. An enduring inscription on a human-readable physical medium may not occur until late in composition process when a relatively complete text is printed to paper.

However, as Matthew Kirschenbaum notes in his excellent and award-winning book, *Mechanisms: New Media and the Forensic Imagination*, there are earlier, encoded inscriptions that happen earlier on storage media that are no less physical despite the fact that they are “invisible to human eyes.”¹

The first scholarship on digital manuscript collections has only started to emerge over the last decade. Christine A. Finn’s journalistic 2002 memoir, *Artifacts*, was one of the earliest works to suggest the then emerging possibilities of digital manuscript studies.² In a chapter that examines the art of Silicon Valley, she describes watching Adobe employee, Julieanne Kost, working in Photoshop:

She paints with layers of images, piling them up, stripping them back—partially, fully, adding again. What she ends up with is distinctive and beautiful, and in a manner of its process, archaeological. [...] The technology allows each of Julieanne’s movements to be recorded in an archive.³

¹ Kirschenbaum, Matthew G. *Mechanisms: New Media and the Forensic Imagination*. MIT Press, 2008) p. 29.

² Finn, Christine. *Artifacts: An Archaeologist’s Year in Silicon Valley*. MIT Press, 2002.

³ *Ibid.*, 40

Finn, an archaeologist, recognized very early on the potential of Photoshop to create artistic palimpsests that reveal the *process* that generates final product.

In a later chapter, Finn briefly examines a print out of the code for the open source Unix text editor, EMACS. In the circa 1982 version of the source code she excerpts, the opening section of programmer comments are written in a combination of Latin and English. These comments are the part of the computer program that is ignored by the processor, but, especially in public open source software, often serve as notes describing the purpose of the code. In the example Finn cites, however, the English comments are terse descriptions of the latest changes to the code (very similar to the “commit” notes submitted when modifications are uploaded to public open source code repositories like GitHub). One of Finn’s technical consultants, Sellam Ismail, a programmer and computer historian who helped to found the Vintage Computer Festival in Livermore, California, tells her, “These comments are too cryptic. I guess they were only meant to be understood by programmers actually working on the code.”⁴ It is, perhaps, surprising that the English comments should be more cryptic than the Latin ones, but both were clearly intended for an audience different than the final “users” of the text editor.

⁴ Ibid., 54

In Finn’s work another notion of digital manuscript emerges—in an age in which art and literature might include programs executed by a machine, the source code behind the program might be considered a manuscript. The public receives a “compiled” executable program on an optical disc (e.g., a CD-ROM) or as a download from an online “app store.” Outside of open source projects, such source code often remains solely in the possession of the programmers who developed on it, and so is usually “unpublished.”

The academic study of these source code “manuscripts” has started to formalize in a field described by some of its practitioners as “Critical Code Studies.” Dennis Jerz’s 2007 recovery and analysis of the source code for the early computer game, *Adventure*, stands out as an excellent early example⁵. More recently, many conference papers and online discussions have been collected on a shared blog dedicated to the field,⁶ and book length studies of digital manuscripts include *10 Print CHR\$(205.5+RND(1)): Goto 10*⁷ (an anthology of essays on a single line of code that creates an animated

⁵ Jerz, Dennis G. “Somewhere Nearby Is Colossal Cave: Examining Will Crowther’s Original ‘Adventure’ in Code and in Kentucky” 1, no. 2 (2007).
<http://www.digitalhumanities.org/dhq/vol/001/2/000009/000009.html>.

⁶ “Critical Code Studies.” Accessed December 26, 2013. <http://criticalcodestudies.com/wordpress/>.

⁷ Baudoin, Patsy, John Bell, Ian Bogost, Jeremy Douglass, and Noah Vawter. *10 Print CHR\$(205.5+RND(1)): Goto 10*. University Press Group Limited, 2012.

graphical pattern on the Commodore 64), and (in the related field of Platform Studies)⁸ Nick Montfort and Ian Bogost's close reading of the technology behind six Atari VCS games: *Racing the Beam: The Atari Video Computer System*.⁹

It would be a mistake, though, to limit digital manuscript studies only to works intended to be experienced on a computer. For at least 20 years, the word processor has been the tool of choice for writers, and so most books written in the 21st century are “born digital” even if they were intended for distribution on paper. Relatively little work has been done on these born digital manuscripts, however, in part because most remain, at present, in the possession of their creators (the majority of the first generation of born digital writers have not yet donated their “papers” to public archives). However, a trickle of floppies and hard drives has already started to flow in libraries and archives around the world. At New York Public Library, for instance, the Lore Segal, Meredith Monk, and Timothy Leary collections are among several that contain digital artifacts. Emory University now serves simulations of Salman Rushdie's early laptops to

⁸ Platform studies is the name Montfort and Bogost gave to their MIT book series that examines the computer hardware that *runs* the code. However, the close connection between code and hardware for the Atari VCS platform makes the margin between this book and Critical Code Studies very thin. While Montfort and Bogost do not reprint any source code, they do describe the coding decisions the early game developers made to realize their vision within the limitations of the platform.

⁹ Montfort, Nick, and Ian Bogost. *Racing the Beam*. MIT Press, 2009.

scholars in their reading rooms. As the writers of the computer era begin to consider their legacy, this flow of encoded bits will soon become a flood.

The amount of information about the creative process scholars will be able to extract from such archives could revolutionize our understanding of how art is made. I have published elsewhere^{10,11} on my research using the 200 floppy disks collected with the Jonathan Larson “papers” at the Library of Congress. These disks document the composer and librettist’s work from 1989 to 1996 and preserve a several dozen versions of both the text and music of *RENT*. When opened in a emulator (a computer program that allows a newer computer to simulate an older one) using a copy of Microsoft Word 5.1 (the word processor Larson used), these files reveal the texts encoded in their original, blocky, 1990s fonts, more-or-less just as Larson saw it when last he clicked “save.”

In all, there are several dozen files on Larson’s disks containing drafts of the script for *RENT*. Together, these files reveal a remarkably complete biography of the text as it develops from a summary with a few song lyrics to the show presented off-Broadway just before Larson’s death. Not all of

¹⁰ Reside, Doug. ““LAST MODIFIED JANUARY 1996’: THE DIGITAL HISTORY OF *RENT*.” *Theatre Survey* 52, no. 02 (2011): 335–340.

¹¹ Kirschenbaum, Matthew G., and Doug Reside. “Tracking the Changes.” In *The Cambridge Companion to Textual Scholarship*, edited by Neil Fraistat, Julia Flanders, Neil Fraistat, and Julia Flanders, 257–273. Cambridge: Cambridge University Press.

the drafts are visible on the “surface” of the files, however. Using a text editor (such as Notepad++ for Windows or TextWrangler for the Macintosh), it is possible to open these same files on a modern operating system and uncover layers of work saved prior to the last save. Microsoft Word 5.1 included feature called “fast save.” In 1992, when most personal computers ran only about 1% as fast as an iPhone 5 and disk drives depended on slow mechanical processes, saving an entire file to a floppy disk was a potentially time consuming activity. The “fast save” feature sped the process somewhat by replacing the entire file only once every 14 saves or so. In most cases, the software would simply append changes to the end of a file with information about where they belonged in the original document. When the file was opened in Word 5.1, the software would integrate these changes back into main text; however, by opening this file with a simple text editor, the original (“base”) text could be recovered and with all of the later emendations revealed at the end of the document. The result is that each of the dozens of individual Word files contain within themselves around a dozen versions of the text.

This *kind* of work, even if the particulars have changed somewhat, is at least analogous to work of scholars of physical manuscripts. The disks are physical objects--once owned by the creator--on which is text is

encoded. Like a physical palimpsest, the disk contains information about the story of the genesis of a text that is obscured in the published copy (even if the publication is digital, like, for instance, the Kindle version of the *RENT* script). However, the current trend towards authoring and storing digital documents in “the cloud” (in services like Google Drive) may change the work of researchers working with writers and composers working today. What are archivists and curators to do when the digital “original” is stored on machines that the author did not own and likely never saw? Will such archives even be accessible after the creators’ death?

Archival staff can encourage living artists to make arrangements to have their digital properties transferred to an archive’s stewardship as part of the acquisition. Already, many archives discuss the acquisition of email with donors, but it is wise to include a discussion of a donor’s full Internet presence. At present, most Google properties, and many other cloud services, have tools that allow users to download all of their content. Using such an export could be part of the acquisition of an artist’s “papers.” There will, of course, be many cases in which collections come to archives after the digital drafts have already been rendered inaccessible (whether due to the unexpected death of the creator, or because the collection was acquired by an archive long after the cloud accounts had been closed or forgotten). This

is part of the loss with which all archivists must make their peace (two things consume and can never be satisfied: the estate executor's bonfire and Google's automatic account deletion scripts). However, archivists can work with the artistic community to encourage creators to make plans for their digital legacy so that even young writers may consider plans for the future of their digital effects.

The value of these digital drafts to a researcher is inestimable, yet can an export of a Google Doc (even with all of its revision history and metadata) be truly considered a manuscript? It is, after all, not the original, but a copy downloaded from a server. In fact, the "original" (that is, the first material transcription of the information) may nothing more than a set of textual fragments stored in a database that is distributed over several geographically disparate machines. A revision comment might first be recorded on a server in Virginia while all of the metadata about the file might first be incarnated in New Mexico. The writer encounters all of this information reconstructed as an integral document, and network efficiencies likely (though not necessarily) ensure that most of the data called up at any one time is likely drawn from the same servers. However, the *original* material object is, less incarnated than *pancarnated*. The materiality of the digital text has not been eliminated, of course, but rather than finding

substantiation on a magnetic wafer might now be almost instantly copied many times from machine to machine in ways likely all but untraceable even to most employees on the Google Drive team.

If digital manuscripts are so massively reproduced even at the moment of creation, what is the different between a manuscript and a published work? Perhaps a digital manuscript becomes a digital publication when it is shared publicly so that those who have no connection to the author can access them? Of course, a manuscript archive then alters the manuscript status of their collection simply by allowing the public to view them in the reading room. Would an archive ontologically destroy a digital manuscript by sharing it publicly online?

This seems unsatisfactory. Perhaps a file is a manuscript until it is modified by a third party (say an editor or a publisher) in preparation for public dissemination? This seems to fit the current usage in the “submission guidelines” for this journal. This essay transfigured from a “manuscript” to a higher form at some moment between the point at which I hit the “submit” button and the point that postal service picked it up from the shipping rooms. Yet, in the archives, digital curators and archivists must frequently perform significant work to make digital manuscripts accessible to modern publics. Are digital libraries, then, publishing houses, changing the manuscript status

of digital collections by preparing them to be served to the public? Are digital manuscript collections simply the equivalent of rare book rooms, presenting (and publishing) manuscripts for very limited runs by processing them through file conversion systems and disseminating them on emulators in readings rooms?

Come to that, what are “rare books” in the digital age? Given the ease with which digital information may be copied and massive distributed, can any digital publication truly be considered rare? Of course copyright and donor restrictions many create artificial scarcity for a time, but if the institutions that maintain these objects are committed to scholarly access, might we assume that there will come a day when most “rare digital books” will become commonly available on the web. Paradoxically, the rarest digital objects may be those most recently created and restricted by copyright and donor restrictions, and rare book librarians those who maintain and mediate access to legally restricted digital information until such restrictions expire.

It is worth noting, though, that if digital rarity is defined as information with very few points of access, then in fact many, perhaps even most, digital works are, in a sense rare. As we are reminded every time a major website is retired and each time Google or Yahoo experiences a

service interruption, in the era of the web, there is a huge amount of important digital information upon which we depend, do not own, but, in a sense borrow from remote servers.

Jason Scott's intrepid team of amateur archivists¹² spring into action when a major web property announces its closure¹³, and the data is stored, for now, on a variety of archival sites including the Internet Archive¹⁴.

Although anyone may copy almost anything from the Internet Archive's vast holdings, the majority of their collection might be legitimately described as "rare." The hassle and expense of expanding, maintaining, and organizing large data repositories coupled with the ease by which files may be retrieved from remote locations when needed, has meant that most of us are content to trust a handful of intuitions and companies to preserve most of our digital culture. Lots of copies keep stuff safe¹⁵, but if a single source has a sense of

¹² "Archiveteam." Accessed December 26, 2013.
http://archiveteam.org/index.php?title=Main_Page.

¹³ Modine, Austin, and 28th April 2009. "Web 0.2 Archivists Save Geocities from Deletion." Accessed December 26, 2013. http://www.theregister.co.uk/2009/04/28/geocities_preservation/.

¹⁴ "Internet Archive: Digital Library of Free Books, Movies, Music & Wayback Machine." Accessed December 26, 2013. <https://archive.org/>.

¹⁵ Reich, Vicky, and David S. H. Rosenthal. "LOCKSS (Lots of Copies Keep Stuff Safe)." *New Review of Academic Librarianship* 6, no. 1 (2000): 155–161. doi:10.1080/13614530009516806.

reliability, and new copies are made very cheaply and easily (as is the case with most web content), there is little incentive to keep the copies we create.

So what are the digital rare book and manuscript collections of the future, and why should society continue to support them? First, digital rare book archives, like the Internet Archive, may be those that provide long-term preservation and access to the otherwise ephemeral content we store in the commercial cloud. But no archive, not even The Internet Archive, can preserve everything. Rare book and manuscript institutions should begin to think about long-term preservation of those cultural objects we now take for granted.

Manuscript and rare book collections may also be centers that house the equipment and employ the expertise necessary to recover and migrate data from old equipment. Platitudes about staff being the most valuable asset in an organization are certainly true in centers such as these. Digital libraries and archives need to employ archivists and curators who can select, operate, and perhaps even design electronic devices to read old media formats and safely preserve bits. Archives have long valued paleographers, diplomatic scholars, and book historians for their ability to make sense of analog data storage devices such as books and papyri, and the new archive must establish collaborations with digital archaeologists and other scholars

who can make sense of and perhaps even describe the information encoded on the bits in our collections. We need now to invest in hiring and training experts in the digital incunables of floppy disks and early commercial software.

If we in archives cling only to the papery past, we may find, as others have suggested, that this first century of commercial computing will be documented in only the most fragmentary way. But for now we have a moment, albeit a brief one, in which many of those who designed these first technology are still alive, and some who know how to maintain them are still working. But we must see born digital preservation not as a supplemental or R&D wing of our archives, but as our now and future core services. Otherwise we risk betraying by negligence those like *RENT*'s Roger and his creator Jonathan Larson who committed to us "One Song to Leave Behind"