

UC Office of the President

iPRES 2009: the Sixth International Conference on Preservation of Digital Objects

Title

Digital Materiality: Preserving Access to Computers as Complete Environments

Permalink

<https://escholarship.org/uc/item/7d3465vg>

Authors

Kirschenbaum, Matthew

Farr, Erika L.

Kraus, Kari M.

et al.

Publication Date

2009-10-05

Supplemental Material

<https://escholarship.org/uc/item/7d3465vg#supplemental>

Peer reviewed

iPRES 2009

THE SIXTH INTERNATIONAL CONFERENCE ON THE PRESERVATION OF DIGITAL OBJECTS

Proceedings

October 5-6, 2009
Mission Bay Conference Center
San Francisco, California



California Digital Library

Digital Materiality: Preserving Access to Computers as Complete Environments

Matthew G. Kirschenbaum¹, Erika L. Farr², Kari M. Kraus¹, Naomi Nelson², Catherine Stollar Peters³, Gabriela Redwine⁴, Doug Reside¹

¹ University of Maryland, ² Emory University, ³ University at Albany, ⁴ University of Texas at Austin

MITH
McKeldin Library
University of Maryland
College Park, MD 20742

Robert W. Woodruff
Library
Emory University
Atlanta, GA 30322

Department of Information
Studies
University at Albany
Albany, NY 12230

Harry Ransom Center
The University of Texas at
Austin
Austin, TX 78713

mgk@umd.edu, elfarr@emory.edu, karimkraus@gmail.com, naomi.nelson@emory.edu, cstollarpeters@gmail.com, gredwine@mail.utexas.edu, dougreside@gmail.com

Abstract

This paper addresses a particular domain within the sphere of activity that is coming to be known as personal digital papers or personal digital archives. We are concerned with contemporary writers of belles-lettres (fiction, poetry, and drama), and the implications of the shift toward word processing and other forms of electronic text production for the future of the cultural record, in particular literary scholarship. The urgency of this topic is evidenced by the recent deaths of several high-profile authors, including David Foster Wallace and John Updike, both of whom are known to have left behind electronic records containing unpublished and incomplete work alongside of their more traditional manuscript materials. We argue that literary and other creatively-oriented originators offer unique challenges for the preservation enterprise, since the complete digital context for individual records is often of paramount importance—what Richard Ovenden, in a helpful phrase (in conversation) has termed “the digital materiality of digital culture.” We will therefore discuss preservation and access scenarios that account for the computer as a complete artifact and digital environment, drawing on examples from the born-digital materials in literary collections at Emory University, the Harry Ransom Center at The University of Texas at Austin, and the University of Maryland.

1. Introduction

Writing is a material act; textual production in any medium has always been a part and product of particular technologies of inscription and duplication. Specialists in the history of the book and other forms of textual studies have long been sensitive to this perspective, and there has been no shortage of significant scholarship attentive to the material qualities of the scripted or printed word and its attendant artifacts, like the codex. Computers, of course, are also writing technologies. Since the popularization of word processing in the early 1980s they have arguably been the dominant writing technology in every segment of society, transforming individuals’ relationships to the act

of composing, editing, and circulating text. Certainly this has been no less true for creative writers of belles-lettres (fiction, poetry, and drama), with prominent early adopters ranging from Stephen King to Joan Didion and Salman Rushdie.

While some writers still continue to work by longhand even to this day, more and more fiction, poetry, and drama is now *born-digital* in the sense that keying the text into a computer (probably relatively early on in its composition), almost always to be further revised, is an all but inevitable part of a contemporary text’s life cycle. Editors edit electronically, inserting suggestions and emendations and emailing the file back to the author. Publishers use electronic typesetting and layout tools, and only at the very end of this process is the electronic text of the manuscript (by now the object of numerous transmissions and transformations) printed as a physical book. In the particular realm of literary and textual scholarship, this means that a writer working today will not and *cannot* be studied in the future in the same way as writers of the past, because the basic material evidence of their authorial activity—manuscripts and drafts, working notes, correspondence, journals—is, like all textual production, increasingly migrating to the electronic realm. David Foster Wallace, to take just one recent example, left behind large portions of an unfinished novel whose manuscript exists both in hard copy and on various computer file systems (Max 2009). If and when this material is edited and published posthumously, the person who undertakes the task will have to be as well versed in legacy storage formats and the idiosyncrasies of Wallace’s electronic writing practices as he or she is in his handwriting and analog composition habits.

The particular case of poets, fiction writers, and dramatists is a specific manifestation of a larger domain that is coming to be known as personal digital papers or personal archives (Cunningham 1994). Typically this

involves the receipt of physical hardware and storage media as part of a hybrid collection blending traditional paper-based materials with either entire computers or computer storage media. Here the challenges and problems begin almost literally on the threshold of the collection's doorstep. What, after all, is being collected? The physical hardware and storage media, or the binary data it contains? How is an archive to contend with hardware and devices that have been in someone's attic or basement for decades? What about the volatility of storage media? (5 ¼-inch diskettes, which were introduced in the late 1970s, have already exceeded their estimated life-span.) Even assuming data can be recovered from these media, can it be authenticated? Stabilized? Does the data consist merely of complete files, or of all the bits contained on the physical media, including perhaps fragments of overwritten or deleted files? How is this material to be cataloged? A single diskette might contain hundreds of individual files, meaning that manual item-level description is prohibitive. A single hard drive will almost certainly contain many *thousands* of files of all types. How is the archivist to know what belongs to the author as opposed (for instance) to a family member using the same computer? What about systems files? What about third-party software included in the author's collection? How can researchers be given access to this material? How will the archivist ensure confidentiality, and that sensitive electronic records will not be copied and redistributed indiscriminately?

While these issues will undoubtedly be familiar to anyone following professional discussions in digital preservation, we believe that literary authors and other creatively-minded originators offer unique challenges beyond those presented by personal digital papers or electronic records originating in domains such as government or commerce.

2. The Site Visits

In 2008 the authors of this paper received funding from the National Endowment for the Humanities' Office of Digital Humanities in support of a series of site visits and planning meetings for personnel working with the born-digital components of three significant collections of literary material: the Michael Joyce Papers (and other collections) at the Harry Ransom Humanities Research Center at The University of Texas at Austin, the Salman Rushdie papers at Emory University's Manuscripts, Archives, and Rare Books Library (MARBL), and the Deena Larsen Collection at the Maryland Institute for Technology in the Humanities (MITH) at the University of Maryland. The meetings and site visits were undertaken with the two-fold objective of exchanging knowledge amongst the still relatively small community of practitioners engaged in such efforts, and facilitating the preparation of a larger collaborative project aimed at

preserving and accessing the born-digital documents and records of contemporary authorship (Kirschenbaum et al. 2009).

Notable authors represented with at least some born-digital material in the collections at either the Ransom Center or Emory (the two major institutional repositories involved in the current research) include Russell Banks, Lee Blessing, John Crowley, Robert De Niro, Michael Joyce, Thomas Kinsella, Bernard Kops, Norman Mailer, Terrence McNally, Tim O'Brien, Salman Rushdie, Ronald Sukenick, Leon Uris, Alice Walker, and Arnold Wesker. Additional creators whose materials the researchers were also able to access include prolific experimental hypertext author Deena Larsen (whose collection is now housed at the University of Maryland) and Jonathan Larson (best known as the composer of the popular musical *RENT*, and no relation to Deena), whose papers (and diskettes) are at the Library of Congress. The sections below elaborate on aspects of these institutional settings and detail the records processing performed to date on select born-digital collections from significant writers.

The Harry Ransom Center at The University of Texas at Austin

The Harry Ransom Center (HRC) is a humanities research library whose primary emphasis is the study of the literature and culture of the United States, Great Britain, and France. In addition to its extensive manuscript, book, photograph, art, and film holdings, the Center also houses the computers and disks of authors such as Michael Joyce, Norman Mailer, Terrence McNally, and Arnold Wesker. The Ransom Center has been receiving born-digital items as part of paper collections for nearly 20 years; as of this writing, thirty-nine of the Center's holdings contain electronic records. These materials include correspondence and manuscript files on a variety of disks and computers. The Center's 2005 acquisition of the Michael Joyce Papers, which, like many recent acquisitions, is actually a digital-analog hybrid collection, marked the Center's first deliberate engagement with born-digital literary materials published in electronic format.

Beginning in 2005, the Ransom Center collaborated with Dr. Patricia Galloway and her graduate students in the School of Information (iSchool) at the University of Texas at Austin to process the born-digital component of several digital-analog hybrid collections. Processing projects completed since 2005 include a pilot project with the Michael Joyce disks, as well as cataloging work on the born-digital materials in the Leon Uris, John Crowley, Arnold Wesker, Norman Mailer, and Terrence McNally holdings. Until recently, the Center housed copies of its born-digital materials in a DSpace repository hosted by the iSchool.

The Ransom Center offers access to these processed collection materials on a case-by-case basis in the reading

room. Between 2006 and September 2009, four patrons requested access, and the Center was able to accommodate all of them. Two of the patrons used the Michael Joyce and Arnold Wesker materials, respectively, via DSpace in the reading room. The third patron accessed copies of files from Terrence McNally's disks from a secure laptop in the reading room, and the fourth patron used a similar set-up to work with copies of Norman Mailer's correspondence files from the early 1990s. In addition, the Center's electronic records collection has been represented in two in-house exhibitions: *Technologies of Writing* (2006) and *The Mystique of the Archive* (2008).

The Center's digital preservation work so far has been markedly collaborative and owes a heavy debt to the assistance of University of Texas graduate students who have processed born-digital collection materials as part of their class projects. The Ransom Center's other productive collaborative relationship has been with the Maryland Institute for Technology in the Humanities (MITH) and Emory University on the NEH project described above. As part of that project, representatives from MITH and Emory visited Austin in November 2008 for a site visit organized by Gabriela Redwine, who is the current digital archivist at the Center, and her predecessor, Catherine Stollar Peters. To give participants a sense of the promise and challenge of the Center's digital collection materials, Redwine and Peters created two different exhibitions that meeting participants were able to access throughout the day. The first consisted of a set of electronic collection materials installed on three computer workstations around the meeting room. The point of these workstations was to give people an idea of what a patron would experience upon visiting the Ransom Center's reading room to look at born-digital manuscripts and correspondence.

Attendees looked at files from four different collections and accessed them from both the desktop and through DSpace. These files included different versions of some of Michael Joyce's hypertext manuscripts, as well as born-digital materials from the Terrence McNally, Arnold Wesker, and Tom Zigal holdings. One of these items, which highlights the intersection of authorship and technology and the palpable influence of technology on an author's work, is a stream-of-consciousness document McNally typed on 10 June 1988 as he experimented with WordPerfect for the first time. Also included was a set of proofs, created in Microsoft Word, that Tom Zigal exchanged with his editor at The Toby Press. Their tracked changes and comments provide valuable insight into the creative process. Both sets of materials offer a precise illustration of the complex motivations for this grant: to understand and preserve authors' works and the environment in which they are created.

The second display was a small exhibition of disks and computers from the Center's collection. Peters and Redwine incorporated these items and their respective histories into an introductory overview to give attendees a

sense of some of the particular digital preservation challenges presented by the media formats in the Ransom Center's collection. In addition, items like a handwritten letter from author Bernard Kops in response to Redwine's query about his computer usage illustrated the importance and utility of beginning conversations with living authors about their technology habits.

One of the most important outcomes from this meeting was a growing awareness of the difference between scholarly and archival perspectives when it comes to thinking about how best to manage, represent, and provide access to born-digital collection materials. A second result was a better understanding of how the Ransom Center's preservation and access strategies compare with those of archivists at other repositories in both the U.S. and England. At the time of the Austin meeting, the Ransom Center had focused preservation efforts purely at the file and series levels and had undertaken little research into preserving disk images. Since then, archivists at the Center have been experimenting with capturing images of disks and hard drives rather than copying individual files directly, and plan to move forward with this methodology as a more comprehensive and less invasive way to capture information from digital media. This is but one way in which the Center's digital preservation program has been influenced by collaborative projects with other institutions and repositories. Finally, a third, more local outcome has been important first steps toward the development of a University-wide community around problems of digital preservation. Archivists from the Ransom Center, the Alexander Architectural Archive, the Benson Latin American Collection, the Center for American History, and the Tarlton Law Library have begun meeting once a semester to discuss the digital preservation challenges at each repository and share information about possible solutions.

Emory University Libraries

The Emory University Libraries (EUL) emerging Born-Digital Archives (BoDA) program has developed as a fundamentally collaborative and strategically innovative enterprise. The team pursuing born-digital work at Emory consists of staff from the Manuscript, Archives, and Rare Book Library (MARBL) and from Digital Systems, representing a range of expertise from archival science and practice to software engineering to digital libraries. MARBL's 2006 acquisition of Salman Rushdie's personal papers provided the Emory Libraries with a rich personal archive of historical and literary significance that includes analog and digital artifacts.

The Rushdie archive marks MARBL's first acquisition of a significant amount of born-digital material and includes four personal computers and one external hard drive. The relationships developed and information shared during the NEH planning grant with project partners MITH

and the HRC proved invaluable to the staff engaged in born-digital archives at Emory. At the start of this planning grant, the BoDA team had completed important preliminary work, such as developing a preliminary project plan for undertaking the handling of the digital material and exploring approaches to organizing and presenting these materials and their analog counterparts as a seamless hybrid archive to researchers. In addition, archivists completed the arrangement and description of Rushdie's analog records in February 2009, while the Technical Lead in digital systems had created masters of each disk image and indexed all five hard drives. Throughout the grant period, BoDA staff continued identifying duplicate files among the machines and assessing the born-digital content. Before outlining more recent progress made on the born-digital materials included in Rushdie's papers, it seems appropriate to first highlight the outcomes of Emory's involvement in the NEH planning grant with MITH and HRC and discuss the impact this partnership has had on the developing born-digital archives program.

As has been discussed earlier in this paper, each institution shared details about its born-digital archival holdings and the current relevant practices and policies. This process of information sharing provided invaluable insights into the range of possible decisions institutions could make about born-digital content. It also gave BoDA team members perspective in which to understand the decisions they had already made about the Rushdie materials and better prepared them for the many decisions they have had to make in the ensuing months since the grant ended. Another significant lesson learned from the grant-funded meetings involved the rich context and insight that can be gleaned from conversations with content creators. At both MITH and Emory, writers joined the grant partners for candid, deeply informative conversations. During the group's interactions with Natasha Trethewey in Atlanta, she described how the use of a word processor encouraged her to experiment more with the arrangement of her words on a page. These conversations not only provided the group with concrete details about how a select few writers interact with technology and understand their digital lives, but highlighted the importance of continuing these conversations. The Emory partners walked away from these discussions committed to building such dialogues into their born-digital archives program.

A final lesson the Emory partners learned while partnering with MITH and the HRC on this planning grant is the value of effective collaboration. The opportunities and challenges inherent to born-digital archives necessitate a community-based approach to developing standards, best practices, policy, and resources. The partnership between MITH, Emory, and HRC quickly grew to include representatives from Yale, the British Library, the Bodleian, and the Georgia Tech Research Institute. Because so many questions remain unanswered those

involved in this field are anxious to confer, to share, to contribute, and to assist. And, on a smaller scale, the initial partners of the grant represented the kind of diversity needed within any group working on born-digital archives. The NEH group included a range of expertise, which required some thoughtful discussions about differences in traditions, cultures, and values amongst the professional fields represented—scholars, technologists, librarians, and archivists. Understanding the importance of these differences helped the Emory partners better manage the diverse working group pursuing born-digital archives at Emory and illuminated that working *toward* collaboration can be just as important as working collaboratively.

As this planning grant ended, Emory partners learned that they would need to prepare the Rushdie hybrid archive for public release in February 2010. This significantly accelerated schedule meant that Emory must quickly implement many of the theories and musings discussed in the preceding months of planning grant meetings. Initially, the focus was on creating a secure dark archive for the disk images and creating a mechanism that would enable access for archivists to obtain copies of the master content. The BoDA working group next began developing the policies and infrastructure for archival processing, which included review for restriction and redaction and compiling basic metadata about each user-generated file. In tandem with these activities, Digital Systems staff began developing tools and interfaces that would enable effective research access to these processed materials. The Emory team is currently loading processed archival content into the tool prototypes, an exciting moment for this young program. Because of the short window for development and processing, the Emory staff has focused its attentions on only one of Rushdie's computers, his Macintosh Performa 5400. The team has developed prototypes of a searchable database that holds discrete files of all approved content from the Performa and of an emulated environment that replicates the original computing environment. Even with the self-imposed restriction to the one computer, staff at Emory have already discovered a wealth of fascinating content and ample evidence for the importance of providing both file-level access and operating system-level access. For instance, Rushdie's use of stickies in his early Mac not only provides insights into his tendencies to meld the personal and the literary but also reveals interesting details about his computing habits.

Despite significant progress in the past seven months, much exciting work lies ahead for Emory's BoDA program. After the Rushdie opening, MARBL staff will continue processing his born-digital files and Digital Systems will work on the second release of the tools and interfaces. Emory is particularly interested in extending researcher access to all five of sets of born-digital content and enhancing the connections between the paper and born-digital materials that comprise Rushdie's archives. Furthermore, members of BoDA will begin gathering data

about user responses to and expectations of born-digital resources within archival settings. Continuing to forge productive relationships such as the one Emory has been so fortunate to develop with both MITH and HRC will undoubtedly be the key to future success.

Maryland Institute for Technology in the Humanities at the University of Maryland

In May of 2007, the Maryland Institute for Technology in the Humanities (MITH) acquired a substantial collection of vintage hardware, software, and other collectible material from the author and critic Deena Larsen. Unlike the Harry Ransom Center or Emory University Libraries, MITH is neither a library special collections unit nor an archive: it is a working digital humanities center. This brings with it certain obvious limitations, but also unique advantages. Founded in 1999 with the aid of an NEH Challenge grant, MITH is the University of Maryland's hub for the theory and practice of digital humanities, cyberculture, and new media, as well as the institutional home of the international Electronic Literature Organization. MITH is thus conceived precisely as an interface between the scholarly and technical communities, a perspective that we think is essential to the current project. At the same time, MITH's institutional situation, encompassing everything from location and physical security to sustainable integration with library resources, creates challenges for ensuring the safety and longevity of an in-house archive. At present, the physical components of the Larsen collection are housed in dedicated (and locked) display cases in MITH's public conference room. Much of the data has been imaged (copied) from the original media, and is stored on a protected server (a so-called "dark archive"). As of this writing, finding aids exist for both physical and digital elements of the collection, and these are in the process of being incorporated as a MySQL database. A public presence for the collection has also been built, featuring a gallery of highlights, access to the finding aids, and background information. But a number of critical tasks remain, chiefly in the realm of item-level description for the digital objects.

While not a household name in wider literary circles, Larsen has been an active member of the creative electronic writing community since its inception in the mid-1980s. She is an avid collector and amateur archivist (or hoarder) who was happy to find a home for the dozen or so vintage Mac Classics, roughly 1000 diskettes, and boxes of journals, papers, correspondence, newspaper clippings, memorabilia, and ephemera previously stored in her apartment. In addition to her own writing and creative work, Larsen also possesses a broad array of material by other electronic literature authors, some of it unpublished, unavailable, or believed otherwise lost. MITH, for its part, looks upon its acquisition of the Larsen collection as both

an important service to electronic literature (by safeguarding what Larsen herself has described as that community's "great library of Alexandria") as well as an invaluable research opportunity, given the potential of this material to function as a testbed.

Larsen's most significant work, *Marble Springs* (1993), exists in a number of physical and digital states which exhibit complex relationships and dependencies. A shower curtain, for example, is the support for a dozen laminated screenshots representing different pieces of the work; these are connected by colored yarn mapping their links and relations. An artifact such as this, coupled with hard copy printouts and transcripts, coupled with digital drafts in various formats and versions of the HyperCard software used as the final authoring environment, is emblematic of the kind of challenge archivists in a number of different cultural heritage sectors can expect to face in the future: not just born-digital content, but digital-analog hybrids. Larsen herself, as a creator, was obviously acutely conscious of the materiality of her electronic medium, embracing not just interface and screen but the whole of the computer as an integral element of the work. The collection at Maryland includes a hand-made "cozy" designed to be placed like a hood over a standard Mac Classic, with openings for screen and disk drive. Moreover, during public installations, the computer running *Marble Springs* (with cozy) was installed on an antique wooden school desk, where the user would sit as he or she perused the work. Larsen therefore imagined a full ergonomics for the end-user's encounter with her work, and designed a hybrid digital/physical space to support its presentation.

Also in 2007, MITH's Doug Reside was invited to inspect and help preserve the digital "papers" of composer and playwright Jonathan Larson (no relation to Deena) at the Library of Congress. Larson is best known as the lyricist and composer of *RENT*. Despite his tragically abbreviated career (he died at 36 from a congenital heart problem), his output was extensive. His papers, including over 150 3 1/2-inch diskettes, were given to the Library of Congress in 2003. Initially the Library had planned to treat Larson's computer diskettes much like the other media in its audio collections—that is, catalog them according to the label on the object but without detailed listings of the files stored on them. However Reside, who is a scholar of musical theater as well as a digital practitioner, suggested an alternative course of action, which was approved by the Library's administration and the Larson estate. It was agreed that he would create disk images (that is, exact, bit-for-bit copies) of the disks in the collection using the "data definition" ("dd") imaging utility that is included with most distributions of the Linux operating system. After creating these "images," he could open virtual versions of the disks on his laptop without needing to work with the actual disks (which would have posed obvious risks). The data images themselves would be stored on a USB flash drive kept in the music library, ensuring that the digital

data, like the physical artifacts, would remain in house at all times.

3. Digital Materiality

Born-digital preservation and records management are still very young specializations. While some impressive guides to best practice already exist (notably the Paradigm Workbook on Digital Private Papers prepared by staff members at the Bodleian and Rylands [Manchester] libraries), and while research is under way in certain quarters, it is clear that the field will remain in a state of flux for the foreseeable future (Johns 2008). Many challenges exist for which there is simply not enough accumulated wisdom and experience to formulate best practices. As we have seen, it is difficult even to achieve consensus on the proper object of preservation.

However, all of this paper's authors share a keen appreciation for what Richard Ovenden has helpfully called (in conversation) "the digital materiality of digital culture." We would gloss this as a curatorial sensitivity toward the uniqueness of individual instances of both hardware and data objects, coupled with an awareness of how the affordances of particular systems, environments, and technologies can all impact the creative process. For example, knowing how much of a document would be visible on a screen at one time—knowledge that depends on the physical size of the display hardware, its screen resolution, and preferences as defined within particular application software—can be critical to understanding aspects of an author's composition process. Terrence McNally comments on precisely this phenomenon in the stream of consciousness WordPerfect document mentioned above. "This is the 22nd line," he writes. "After I finish it and two more, the screen should begin to move upwards and I will only be seeing the last 25 lines. It is not possible to see an entire document when you work with a computer." Umberto Eco had Belbo similarly experiment with his new computer in the novel *Foucault's Pendulum*, and Salman Rushdie has equivalent files on his Macintosh laptops, showing that he, too, took time to explore the environment of his new computer. The experience of an author composing on a Mac Classic from 1985 will be different from the experience of an author working on a contemporary wide-screen LCD display, or perhaps several such displays configured in tandem. It is easy to forget that even a mundane task like erasing a block of text has changed dramatically since the earliest days of personal computing. For example, the journalist James Fallows, writing about his first word processor (a Processor Technology SOL-20) in 1982, describes how he must place special marker characters at the beginning and end of the passage to be removed, then execute a series of chorded keystrokes to delete it (Fallows 1982). Textual scholars have been attentive to the "materiality" of books and manuscripts for decades, especially following the

intervention of scholars such as D. F. McKenzie and Jerome McGann who formulated influential approaches to the theory of scholarly editing in the 1980s (McKenzie 1986, McGann 1991). Heather MacNeil, meanwhile, discusses correspondences between the textual scholar and the role of the archivist (MacNeil 2005). As we move forward into the digital era, we would do well to remain attentive to the material conditions of computing, and the way in which these material conditions, often as much socially determined as purely technological, contribute to the end-user experience.

Closely related to these questions of materiality is the hybrid status of nearly all born-digital collections of which we are aware, in which electronic objects coexist with more traditional forms of archival content. Material from collections at all three of our institutions exemplifies this phenomenon, with the textual horizons of a particular work often spanning multiple media and formats, from holograph manuscript to hard copy print-out of a born-digital text, to actual digital files. (The Larsen shower curtain at Maryland is perhaps the limit case.) Scholars will want and need to track the evolution of a work without regard for the gaps and incompatibilities introduced by competing or obsolescent data formats and operating systems, let alone the analog/digital divide. Yet there are no tools to facilitate this kind of activity, and there are unlikely to be for the foreseeable future. Compounding the problem is the reality that working authors often gravitate toward proprietary software, such as the word processor that came installed with their system as a default. While various communities have had reasonable success to date in developing text analysis, text mining, and visualization tools for large electronic corpora, these tools often assume ideal circumstances and a homogeneous data set, not the messy world of proprietary and mutually incompatible formats one gets from an individual user's hard drive. At one end of the spectrum we can therefore anticipate expanding metadata for finding aids to more robustly track the migration of a work across multiple media and formats. At the other, more exotic, end of the spectrum it is perhaps possible to imagine grafting RFID tags to physical archival objects in order to convert them to what Bruce Sterling has called "spimes"—that is physical objects digitally locatable in space and time—thereby making linkages to associated data explicit. Regardless of what solutions are actually deployed, it is clear that both archivists and scholars will need to contend with increasingly complicated ecologies of primary source documents spanning heterogeneous digital and analog states.

The new and formidable challenges presented by cloud computing—that is the increasing reliance of network-centric services for email, blogging, photo-sharing, and social networking—complicate these considerations of materiality, as does the growing user base for third-party back-up services like iDisk and Carbonite (Garfinkel and Cox 2009). It may be, in fact,

that content from the first twenty-five years or so of personal computing represents an anomalous window of opportunity wherein the archivist enjoys reasonable prospects for access to the original hardware and storage media. If that is the case, than distinct preservation strategies suitable for that circumstance are all the more necessary.

Computers are writing technologies, but they are also *environments*: work spaces, surrogate desktops that function as extensions of self. As computers become more and more integrated into our daily routines they become the site for managing multiple aspects of our lives, the windowed screen playing host to a manuscript draft one moment, an email message the next, perhaps a financial statement or a family photograph thereafter. We personalize our computers—and to a large extent we inhabit them. Should a scholar be allowed to see an author's high score on Tetris or their choice of desktop wallpaper? (J. K. Rowling's fans are known to obsess over her scores on the popular game Minesweeper.) What about the music available on an MP3 playlist? What about choices for fonts and layout? Such details may seem trivial, but in fact scholars often want to know what an author was listening to or what images were important to him or her during the writing process. The most mundane features of modern operating systems quickly blur the distinction between the "system" as a generic architecture and the idiosyncrasies of its user-created environment. A computer's registry, for example, stores information related to all of the device drivers and application software in the operating system. Access to the registry is among the most invasive procedures an outsider could undertake; but its value as a record of the digital environment of the computer is enormous. At the level of individual works, scholars will surely want to examine a file's properties, which contain records of when it was last opened and closed and how many hours and minutes was spent accessing it. This kind of metadata, while hardly infallible—it could be spoofed by something as simple as an incorrect system clock—could, with care, be used to establish chronologies that could date the composition of a work—or specific passages within a work—to the hour, minute, and even second. Should a scholar be permitted to cross-reference this kind of information with, say, the downloaded internet files residing in a Web browser cache? And what about pornography or other sensitive material that turns up on the machine?

While it will obviously take some time and experience to balance needs and opportunities for scholars and other patrons with donor privacy and legal restrictions on certain types of information, we believe that in the interim it is important to ensure we are not foreclosing options by inadvertently failing to attend to key elements of the materiality of the original hardware and media as it is accessioned and cataloged.

4. Conclusions and Recommendations

Archivists and other information professionals have long appreciated the impossibility of predicting all potential use-case scenarios for items in their collections. Brown and Duguid recount the episode of the medical historian opening boxes of dusty letters, not to read them but to sniff their envelopes for traces of vinegar (used as a disinfecting agent) in order to reconstruct the course of a cholera outbreak, for example (Brown and Duguid 2000). G. Thomas Tanselle, meanwhile, has long been an advocate against library practices that discard original dust jackets and rebind books on the grounds that bindings and jackets constitute essential evidence for those interested in the history of the publication of the book (endorsements, for example, which are not always duplicated in the interior text, or cover art) (Tanselle 1998). Experience suggests that material or environmental evidence is indispensable when dealing with the artifacts and records of individuals for whom we wish to know as much as possible about their qualities of mind and creative process, as well as the social circumstances surrounding their work. Jane Austen's residence at Chawton still preserves the famous creaking door which would have warned her of a visitor's approach, since at the time novel writing was considered unseemly for a woman of her station. As preservationists we would therefore do well to ask: what are the dust jackets of the digital age? What seemingly incidental features of the digital environment may turn out to have value for a researcher whose future interests we cannot foresee? What software (perhaps even spyware) is the equivalent of Jane Austen's creaking door?

Here then are some basic conclusions and recommendations we have drawn from our study. First, hardware and storage media may themselves possess evidentiary value. At the very least, these can function as numinous objects (as evidenced by their display in exhibitions at the Ransom Center and confirmed anecdotally by the spontaneous response of Kirschenbaum upon being shown a laptop belonging to Michael Joyce). Decals and stickers on a laptop, nicotine or food stains on a keyboard, the label on a disk—all of these are examples of material evidence that might prove of value to a researcher. Therefore, for archivists and others working to preserve born-digital materials, there is a strong argument for preserving the integrity of the original hardware and storage media accessioned with a collection, however generic or unremarkable these might appear. Moreover, as the example of screen size in the previous section illustrates and as video game preservation enthusiasts have long understood, physical hardware components can be essential to understanding the affordances of an obsolescent system.

Second, we are strong proponents of imaging hard drives and other disk media. While resources, including staff time and storage capability can mitigate against large-scale disk imaging, the costs of obtaining and storing complete images of original media are modest compared to the value these materials may yield for future generations

of researchers. A disk image that includes an operating system allows a future user to reconstruct the complete digital context for the originator's work, including the software in which files were created as well as seemingly incidental features such as desktop wallpaper or Preference settings. Moreover, a disk image retains bit-level data that is lost in standard file copying. Disk images raise obvious issues of privacy and data security, and will therefore require appropriate data handling regimens, as well as careful and comprehensive discussions with donors.

Third, we believe that these sorts of collections are ideal candidates for forensic recovery techniques. Here the donor's wishes must obviously remain paramount, but while no responsible archivist would willfully violate a donor agreement, one wonders what would happen if we somehow had access to (say) Shakespeare's hard drive at this point in history. If a contemporary author were to attain comparable cultural stature, who can say what future generations might wish to do if the promise of recovering some lost masterpiece were at stake? Beyond data recovery, forensic techniques are invaluable for stabilizing and authenticating data.

Fourth, we advocate documenting as fully as possible the original physical settings in which the donor's computers were used. This would include photographs, video, and even virtual models of the workspace. Such documentation differs little from what scholars and biographers have been doing for decades, but here the computer must be appreciated for what it is: the nexus of the creative process, not just as a utilitarian appliance.

Fifth, we believe it is essential to talk with practicing writers about their digital work habits. As described above, these conversations during the grant period were invaluable. We were interested in the most mundane details, such as whether composition begins at the keyboard or if they work with pen and paper, how often they save revisions and versions, whether they have a Web browser open while they write, how they handle their email, and whether and how they think about the privacy issues that would arise with a forensic or archival examination of their computer.

Sixth, we believe that user needs and interface requirements for users engaging born-digital literary material are a vital area of future study. How will patrons access born-digital records in a manner that preserves their material integrity yet assures appropriate donor privacy and data security? How might user interfaces for born-digital creative materials differ from other user interfaces?

Seventh, we believe that preserving computers as complete environments means opening appropriate channels of scholarly communication. How does one cite a passage of text in a digital file running in an emulator, for example? (It is often difficult to even copy and paste text from an emulator into a text editor on the same system.) How does one incorporate information about software and versions of files into scholarly citation? Scholars will demand robust modes of engagement with born-digital

materials and efficient pathways between these materials and their own systems of scholarly communication.

References

- Brown, J.S. and Duguid, P. 2000. *The Social Life of Information*. Cambridge, Mass.: Harvard Business School Press.
- Cunningham, A. 1994. The Archival Management of Personal Records in Electronic Form: Some Suggestions. *Archives and Manuscripts* 22.1 (1994): 94-105.
- Eco, U. 1988, 1990. *Foucault's Pendulum*. London, Picador.
- Fallows, J. 1982. Living with a Computer. *The Atlantic* (July).
- Garfinkel, S. and Cox, D. 2009. Finding and Archiving the Internet Footprint. First Digital Lives Research Conference: Personal Digital Archives for the 21st Century. London, England.
- John, J. L. 2008. Adapting Existing Technologies for Digitally Archiving Personal Lives: Digital Forensics, Ancestral Computing, and Evolutionary Perspectives and Tools. In iPRES 2008: The Fifth International Conference on Preservation of Digital Objects. British Library, London.
- Kirschenbaum et al. 2009. Approaches to Managing and Collecting Born-Digital Literary Materials for Scholarly Use. Office of Digital Humanities, National Endowment of the Humanities:
<http://www.neh.gov/ODH/Default.aspx?tabid=111&id=37>
- MacNeil, H. 2005. Picking Our Text: Archival Description, Authenticity, and the Archivist as Editor. *The American Archivist* 68 (Fall/Winter 2005): 264-278.
- Max, D. T. 2009. The Unfinished. *The New Yorker* (March 9, 2009).
- McGann, J. 1991. *The Textual Condition*. Princeton: Princeton University Press.
- McKenzie, D. F. 1986. *Bibliography and the Sociology of Texts*. London, British Library.
- Paradigm Project. 2005-7. *Workbook on Digital Private Papers*. <http://www.paradigm.ac.uk/workbook>.
- Sterling, B. 2005. *Shaping Things*. Cambridge: MIT Press.
- Tanselle, G. T. 1998. Bibliographers and the Library. In *Literature and Artifacts*. Charlottesville, Va: Bibliographical Society of America.
- Terrence McNally Papers, Harry Ransom Humanities Research Center, Disk No.: 22a, File name: NEWLIGHT.