Archive and Database as Metaphor: Theorizing the Historical Record

Marlene Manoff

abstract: Digital media increase the visibility and presence of the past while also reshaping our sense of history. We have extraordinary access to digital versions of books, journals, film, television, music, art and popular culture from earlier eras. New theoretical formulations of database and archive provide ways to think creatively about these changes to the cultural and historical record. This essay explores the ways in which the current digital environment can be theorized in terms of, what I call, its archival effects.

Introduction

Our relation to history is being transformed by digital media that increase the presence of the past and that remake our connection to both past and present. New theorizations of archive and database enable us to think productively about the nature of digital objects and the ways in which they alter cultural memory and historical transmission. This paper analyses the growing theoretical literature on the complex and contradictory effects of digital archives and their implications for the historical record and the future of libraries.

Archive and Database as Metaphor

When scholars outside library and archival science use the word "archive" or when those outside information technology fields use the word "database," they almost always mean something broader and more ambiguous than experts in these fields using those same words. The disciplinary boundaries within which these terms have been contained are eroding. Scholars use the terms metaphorically, appropriating them from the professional experts.[1] Some writers exploring the concepts of archive and database begin with standard dictionary definitions. Archive is defined as "a place in which public records

or historical documents are preserved;"[2] database is defined as a "large collection of data organized esp. for rapid search and retrieval (as by a computer)."[3] But archive and database have also evolved into increasingly contested terms used to theorize digital culture and new forms of collective memory.

My concern here is primarily the way that archival theory has informed our thinking about digital objects and the theorization of the term "database." In an earlier essay, I traced the explosive growth of archival theory outside the fields of library and archival science.[4] In this essay, I explore the tremendous archival power of the digital and the value of archival metaphor in helping us to understand new media. Scholars are exploring the social and political implications of, what I call, the "archival effects" of the digital environment. "Archival effects" is meant to suggest the ways in which digital media bring the past into the present. We have extraordinary access to digital versions of books, journals, papers, film, television, music, art, and popular culture from earlier eras. We inhabit a present in which more people than ever can read, watch, listen to, study and collect digital versions of historical objects. This proliferation of digital media has led critics to explore the social and political implications of the database form. Theorists are declaring database to be the prototypical genre of the twenty-first century and the latest embodiment of both the epic and encyclopedic impulse.[5] Database is the most recent in a progression of forms that support or give voice to the instinct to create, collect, and transmit culture.

Following the work of Lev Manovich, critics have been adopting his notion that "database" is the reigning paradigm for contemporary culture, providing a way to describe and understand the new ways in which we experience reality.[6] Manovich argues that our experience of culture, both past and present, now comes to us filtered through the human-computer interface. This means that, as the use of digital devices expands, the database experience increasingly determines the nature of our connection to knowledge and history. Our access to the archive is becoming more dependent upon the technologies of the interface, even as the interface is being transformed to accommodate a host of new digital devices.

In both scientific and humanistic disciplines, digital technologies have fostered the creation of vast collections of text, images, and data. Archival metaphor provides a way to theorize this exponential growth. Jacques Derrida's formulation of an archival impulse or archive drive[7] underpins a significant portion of this work including, for example, media historian Belinda Barnet's description of the birth of the Internet.[8] Explaining its early development, she traces an impulse to create "a giant repository of ideas, a cultural memory," and "to organize and archive the great body of human thought."[9] We might, therefore, say that our current moment reflects the convergence of two phenomena— new technical capacities and an age-old impulse to gather and preserve. The ease of capturing digital data is an incitement to archive.[10]

The Presence of the Past

Digitization provides both greater access to the past and a reshaping of historical consciousness. While cultural artifacts from earlier periods have been accessible through libraries, archives, and museums throughout the modern period, the ease of access and

the omnipresence of older artifacts now available in digital formats is a recent phe-nomenon. Digitization has enabled the dissemination and teaching of texts that were previously unavailable in print. The Women Writers Project, for example, pioneered the digitization of works by early women writers that were either out of print or that existed only in manuscript form.[11] Similar projects have vastly multiplied the options for reading, researching, and teaching material from earlier periods. This, in turn, has led to the writing of new histories based on newly available primary sources that reshape our notions of the past.

The digitization of out-of-print texts has in some cases led to the print publication of individual works and/or their inclusion in print anthologies. The Women Writers Project collaborated with Oxford University Press to publish 15 separate volumes by women represented in their databases.[12] In addition, they sell paper copies of individual works, making them available to both researchers and students for the first time. A somewhat surprising archival effect of the digital, therefore, has been to increase the availability in print of previously rare historical materials.

The expanding digital universe offers historians access to new kinds of materials that will enable them to conduct new kinds of research. In April 2010, for example, the Library of Congress (LC) announced that Twitter agreed to donate an archive of its public Tweets and would be providing regular updates.[13] Users of Twitter now broadcast about 55 million Tweets per day.[14] LC intends to preserve this archive for posterity. Randall Stross of the *New York Times* describes the Twitter archive as containing "more observa-tions, recorded at the same times by more people, than ever preserved in any medium before."[15] Twitter is also making its Tweets available to Yahoo, Google, and Microsoft.[16] Google allows users to conduct a search and then limit the results to Tweets. These results can then be limited by time period. Because they are in digital form, Tweets are searchable in ways many first-hand historical accounts are not. An archive of Tweets will provide historians with knowledge of historic and other events from across the globe as they unfolded in real time and as they were experienced in the moment by users. These kinds of digital affordances are transforming the nature of historical research.

Digital technologies produce contradictory effects in that they appear to both re-pudiate and recover the past. For some, digital technologies signify a new age that has evolved beyond printed books and libraries. Yet, it can also be argued that the "rela-tive novelty of computing technology itself…disguise[s] the fact that this machine also perpetuates the old."[17] The digital may be seen as a radical break with earlier modes of representation, yet it also fosters continuity through recovery and recirculation of vast numbers of historical texts and artifacts. Richard Lanham is one of many writers who describe the ways in which digital technologies allow us to revisit and reclaim the past.[18] He observes that there is more classical music available now than ever before and at relatively low prices. He notes the current availability of large numbers of art house films through Netflix and other sources. And Lanham particularly celebrates the vast global market for used books provided by Amazon and ABE.[19] Google Books and the Internet Archive make large numbers of out-of-print and difficult-to-obtain volumes available to a broad audience. Further expanding our access to historical knowledge, computer modeling and visualization tools are being used to create virtual reconstruc-tions and databases of ancient objects, structures, and even entire cities.[20] These new

digital affordances serve as partial substitutes for historical sites threatened by war, tourism, urbanization, looting, and decay.[21] More people than ever can now (virtually) visit historic heritage sites. Computer tools and computer memory expand, increasing the weight of the past while the present appears to shrink through accelerating cycles of innovation and obsolescence.

Digital technology creates an appetite as well as a market for the historical objects it delivers and recontextualizes. Increased storage capacity fosters the creation of collections of historical films, images, and music as well as texts. As Will Straw argues, digital culture refashions and refocuses the past.[22] Older works may be packaged as sets with annotation and commentary in ways that serve to naturalize particular interpretations. Digitization provides access to selected artifacts of the past, both scholarly and popular, ordered and contextualized by producers and distributors. The marketing of older film and music transforms and reframes them as part of the current zeitgeist.[23] The release of sequels or prequels to films like *Die Hard, The Terminator, Batman, Toy Story, Shrek,* and *Star Trek* revive interest in earlier installments and thus reinvigorate these franchises as commercial properties.[24] Older films are re-mastered and made available on cable and in new formats like Blu-ray as a way of creating interest in sequels or prequels.

New technologies inspire the creation of hybrid objects that provide new historical contexts. The colorization of black and white film is an example of how older media objects may be introduced in new forms. When Martin Scorsese made *The Aviator*, a film based on the life of Howard Hughes, he colorized and incorporated footage from films made by Hughes.[25] Contemporary access to huge media archives provides the opportunity for displaying, recontextualizing, and commenting upon both current and older news events. Much of Jon Stewart's and Stephen Colbert's best satire involves the juxtaposition of contemporary and/or historical media clips. Our notions of the past are shaped by our exposure to various selections or packagings of historical artifacts, by hybridized versions of older cultural objects, and by new media that incorporate and frame pieces of the past.

New and old media re-circulate in new contexts, annotated, tagged, and/or assembled by scholars, commercial producers, or individuals. In this way, digitization alters our sense of history. Or, to use Henry Jenkins' somewhat different formulation: "If the digital revolution paradigm presumed that new media would displace old media, the emerging convergence paradigm assumes that old and new media will interact in ever more complex ways."[26]

For libraries, this proliferation of historical artifacts in digital form suggests that any impulse to treat their print collections as quaint manifestations of a departed era may be premature.

Digital Media and Historical Continuity

For libraries, this proliferation of historical artifacts in digital form suggests that any impulse to treat their print collections as quaint manifestations of a departed era may be premature. Certainly more books from earlier periods are more visible than ever through search engines and online catalogs. For many, digital access to monographs enhances but does not substitute

for print access. Collections of digital images may inspire viewers to seek out original artifacts in galleries, museums, libraries, and archives. Any simple formula presuming the inevitable replacement of historical artifacts by digitized reproductions ignores the contradictory forces at work in contemporary culture, including renewed scholarly and popular interest in cultural objects as material artifacts. As recently argued by Jerome McGann, the volume and scale of contemporary digitization projects imposes two major responsibilities on educators and scholars: "to oversee and monitor this process of digitization" and "to protect our paper-based inheritance."[27]

Digitization of historical artifacts always involves a process of transformation. A digital text differs from its print (or stone or manuscript) original. As I have argued elsewhere, analog and digital objects, print and electronic texts are shaped by the technologies used to produce and distribute them and, therefore, provide different kinds of experiences for users.[28] The availability of massive historical digital archives is a huge benefit to anyone interested in the past, but digital surrogates should not be confused with originals. Scholars increasingly argue that the meaning of a work, whether print or electronic cannot be separated from its physical manifestation.[29] Roger Chartier, who applauds the benefits of widespread digitization, nevertheless declares that it is "fundamentally wrong" to assume "an equivalence between media and that a text is still the same regardless of its form: printed, microfilmed, or digital."[30]

Globalization and rapid social change induce nostalgia for the past even as the digital environment affords considerable intimacy with artifacts of earlier eras. We are in the odd position of having tremendous access to historical artifacts and digital surrogates while also experiencing a sense of being cut off from history. The current emphasis on the use of primary source material in teaching as well as the popularity of historical and genealogical material online is a function of the widespread availability of digital access to these sources, but it is also a symptom of a desire for historical continuity. Libraries and archives play a key role in maintaining and enabling this continuity.

The Commercialization of the Archive

The Whitman Archive, the Rossetti Archive, the Blake Archive, and the Perseus Project are examples of freely available digital collections providing access to important historical material. They are supported by grant money from the National Endowment for the Humanities (NEH), the Mellon Foundation, and the MacArthur Foundation. All three are helping to fund the creation of infrastructure for digitizing and preserving the historical record.[31] Peter Stallybrass describes the Walt Whitman Archive as one example of the liberating effect of online access to material that was previously available to only a privileged few.[32] The Walt Whitman Archive eventually seeks to make the entirety of Whitman's work as well as translations and secondary material freely available online. Stallybrass applauds this democratized access as it has "profoundly undermined an academic elite's control over the circulation of knowledge."[33] But, unfortunately, the archival impulse operates less beneficently in the corporate sphere. Companies like ProQuest and Gale (now Cengage Learning) own huge portions of the scholarly and historical record. Gale has created massive databases containing digital versions of most English language monographs published prior to 1800. To scholars studying the

eighteenth century, it is a distinct disadvantage to work at an institution that cannot afford to purchase Gale's Eighteenth-Century Collections Online (ECCO), which includes almost every significant English-language title printed in Great Britain and the United States during the eighteenth century. Unless one has access to a library that can afford the hundreds of thousands of dollars required to purchase databases like ECCO and Early English Books Online (EEBO), one may not be inclined to celebrate these particular manifestations of the archival impulse. Digitization does not lead in any simple or straightforward way to the democratization of knowledge.

Archival ambitions are also key to many of Google's enterprises. Besides its massive digitization projects, Google aspires to provide a searchable interface to the entire world of information. One price we pay for the ease of searching Google is a lack of transparency. As William Turkel pointed out in 2008, more than half of the billions of searches performed worldwide each month are handled by Google, "making its ranking algorithm the most pervasive source of bias in the history of research."[34] Turkel is not alone in his concerns about the impact of Google searching algorithms and the need to better understand the search engines that play so large a role in contemporary research. Many scholars worry about the ways in which databases and search engines filter and constrain research queries. Jonathan Freedman, for example, warns against the tendency "to downplay the inclusions, exclusions, choices that have gone into the making of databases" and against the failure to question or understand the results.[35] Scholars in many fields are finding that, even if they have access to scanned images of material provided (for example) by Google or ECCO, the quality of the scanning and accuracy of the search mechanisms may be insufficient for their purposes.[36]

One price we pay for the ease of searching Google is a lack of transparency.

Librarians also worry about the impact of researcher dependence on Google and the outsize role it plays in searching and scanning large portions of the historical and cultural record.[37] Some librarians have been particularly alarmed by the proposed Google Book settlement. In a letter to the U.S. Department of Justice in late 2009, three major library organizations expressed concern that Google's monopolistic control of orphan works would expose libraries to predatory pricing for institutional subscriptions.[38] Corporate control of the digital record is having and will continue to have a tremendous impact on the historical record. Laura Mandell, a scholar of the British romantic period, argues that "digitization of the historical record for commercial reasons distorts it by selecting what is preserved, what is searchable, and how on the basis of profit rather than scholarly value and need."[39]

Corporate control of the digital record is having and will continue to have a tremendous impact on the historical record.

Similar concerns are being voiced outside academia. Jimmy Wales, a co-founder of Wikipedia, points to a growing consensus that it is "unhealthy for the citizens of the

world that so much of our information is controlled by such a small number of players, behind closed doors."[40] Although in some cases the creation of digital collections has made the historical record much more broadly available, in others it has led to corporate ownership of both scholarship and primary source materials on a scale that would have been unthinkable just 20 years ago. Corporate ownership may impose monetary barriers to access as well as dependence upon poorly scanned images and / or inadequate search and retrieval tools. And it may also determine which versions or editions scholars are able to consult, thus creating de facto canons of primary resources.

Among the more positive developments, new creative collaborations in the non-profit sector are contributing to the expansion of freely available digital archives. The National Endowment for the Humanities (NEH) and the German Research Foundation (DFG) are jointly funding digital projects of interest to American and German scholars.[41] The NEH is also partnering with the Joint Systems Committee (JISC) in the United Kingdom to offer grants for collaboration between U.S. and English or Welsh institutions for expanding current digitization projects, creating pilot projects, or developing infrastructure to support digitization.[42]

Public private partnerships may provide an alternative to unfunded or underfunded initiatives in the public sector or to purely commercial digitization projects. The European Commission recently released a "Final Report on Public Private Partnerships for the Digitisation and Online Accessibility of Europe's Cultural Heritage."[43] The report suggests that partnerships between commercial entities and public sector institutions (which include private institutions of higher education) may help with "funding, technology, software and expertise required for large-scale digitization."[44] The report describes nine projects in a variety of languages and disciplines and provides a list of criteria for successful partnerships: maximizing public access, considering long-term preservation and sustainability, operating within applicable copyright law, avoiding exclusive agreements, and operating transparently.[45] Not all of the projects described meet all of the criteria. Commercial entities, public sector institutions, and public-private partnerships will each continue to play a role in digitizing and disseminating our cultural legacy. At this point, it is unclear how much of the historical record will be publicly available to future generations. It is also unclear who will ultimately be empowered or disempowered by the growth of digital archives. Given the large volume of freely available digital material, many may be unaware of the vast amount of commercially owned scholarship and data that are invisible or only partially visible to them.

The Ubiquity of the Archive

Our relation to the archive is being reconfigured by the sheer volume of information emanating from our computers and digital devices. The archive is no longer a collection of artifacts, books, and records confined to particular locations that we may seek out if we so desire. Much of the archival record now consists of streams of data invading our work and private lives, perpetually tempting us to consume or contribute just a little more. The state, large memory institutions, and media companies are no longer the sole superintendents of the archival record. Individuals now have the tools to build their own digital archives and organize and manipulate them. This proliferation of digital media,

wireless technology, and mobile devices has begun to erode the boundaries between virtual and material worlds. Mobile devices foster constant connection to the archive. We are immersed in and no longer separable from the data streams in which we swim.

Writers are heralding an end to cyberspace as we have known it. Alex Pang argues that "[a]s the Internet becomes more pervasive—as it moves off desktops and screen and becomes embedded in things, spaces, and minds—cyberspace will disappear."[46] Or, in Paul Hartzog's phrasing, "As the information world becomes layered onto the physical world by mobility and ubiquity, the whole online/offline distinction becomes less useful as a framing metaphor."[47] Cyberspace, and thus the electronic archive, is being appropriated and transformed into part of our daily material lives. "The real digital archive is not concentrated in formal mechanisms for legal deposit and preservation, but saturates the entire social network. By sharing information, we become part of a living archive."[48]

"Database as Genre"

This vast collection of disparate digital data is useful because databases make it searchable and navigable. Lev Manovich, as noted earlier, has declared database to be the preeminent cultural form of the current moment. He points out that, although strictly speaking they are not databases, new media typically "appear as collections of items on which the user can perform various operations—view, navigate, search."[49] Citing both Manovich and Derrida, Ed Folsom, one of the founders of the Walt Whitman Archive, proposes that database is a new genre that is transforming the nature of archives.[50] In a special issue of *PMLA* devoted to "Remapping Genre," Folsom describes database as a way of transcending some of the physical and financial constraints of a paper archive.[51] He celebrates the increased opportunities afforded by digital as opposed to print technology, and he suggests that many works from earlier periods possess characteristics that make them particularly well suited to digitization. Folsom sees Whitman's *Leaves of Grass*, for example, as "genre-bending," especially the first 1855 edition with its mixture of journalism, oratory, and biblical allusion.[52] Folsom claims that, for Whitman, "the world was a kind of pre-electronic database."[53]

Jonathan Freedman pushes Folsom's metaphor further. He finds in Whitman's work an "encyclopedic impulse," a move to "inventory, name, define, and (partially) order" the world.[54] He calls this Whitman's "will to database…his desire to enumerate and catalog."[55] Various contributors to this special issue of *PMLA* explore the evolution of a genre, beginning with the epic and continuing through the development of encyclopedias, commonplace books and catalogs, through to today's databases. All these genres perform archival functions in that they are aggregators or accumulators of knowledge; they save and transmit culture. These formulations implicitly make the case for historical continuity. They imagine current configurations of database as a kind of logical unfolding of the archival impulse.

Folsom claims that the use of the word "archive" in the name Walt Whitman Archive is primarily metaphoric and that the Whitman archive is more accurately described as a database.[56] Jerome McGann responds that such a characterization is "seriously misleading" and argues for a much narrower definition of database.[57] McGann sees database as

a blunt instrument and suggests we have much work to do if we are to create digital environments and tools that support the full range of humanistic inquiry. McGann believes that collections of texts that have been digitized and encoded in markup languages like SGML or XML should not be considered databases.[58] This is essentially a disagreement about how broadly to define database and the usefulness of database as metaphor.

Database and the Death of Narrative?

But the question of database as metaphor is more vexed. When McGann objects to the metaphoric use of the term "database," he traces the problem back to Lev Manovich's theorization of an opposition between database and narrative forms and the misuse of this theorization by literary scholars. Manovich grants that narrative still persists in some new media objects and that what we are seeing is primarily a change in emphasis from narrative to database modes of cultural perception. But because Manovich insists on characterizing modern media as "the new battlefield for the competition between database and narrative" and because he declares that the "database impulse" is becoming "the logic of culture at large," many see him as pronouncing upon the slow but inexorable death of narrative.[59] This is too reductive for many scholars.

Manovich's case for the decline of narrative is undercut by his own reliance on narrative. Any account of media transformation implicitly requires a narrative structure. Database modes may be proliferating, but narrative remains central to history; cultural memory; national, religious, ethnic, and personal identity; and to a host of disciplines in the humanities, social sciences, and sciences. A more productive way to parse these terms is provided by Katherine Hayles, who describes narrative and database as complementary forms. She proposes that they be seen as "natural symbionts."[60] Hayles claims that the more data we produce, the more we need narrative to make sense of that data and to find constructive ways to use it.[61] Data are only useful insofar as we are allowed to create meaning and thus construct productive applications.

Reconciling Database, Narrative, and Archive

The Orlando Project, a massive history of women's writing in the British Isles, provides another way of thinking about both the database/narrative opposition and the archive/database nexus.[62] The project consists of thousands of segments of text that have been encoded in SGML markup language. The founders of the Orlando Project claim that the digital environment fosters the multiplication of narrative rather than its elimination. They argue that their process of constructing a digital literary history demanded the incorporation of narrative elements that were "multiple, parallel, and fractured rather than continuous and singular."[63] It is precisely this multiplication of narrative that positions the project to counter various limitations of traditional literary history, including its linearity and "totalizing or monologizing tendency."[64] Database is thus conceived as an amplification of narrative, rather than a replacement.

This argument is echoed by Caroline Levander, who advocates an approach to history that moves beyond national frameworks, focusing instead on hemispheric research models.[65] Levander claims that new digital tools allow researchers to break away from "strict,

linear narratives of modern development—be they historical, anthropological, literary, religious, sociological, biological, or economic" that "fail to capture the multidimensional, multidirectional, and palimpsestic nature of hemispheric research."[66] Levander declares that digital archives present researchers with unprecedented opportunities to "reconceive the organizing premises of stored knowledge and to make hidden texts, material, and pasts immediately apparent."[67] This too suggests that digital archives are contributing more to the multiplication of narrative than to the death of narrative.

Although one might describe the Orlando Project as a digital archive, its creators decided not to use that terminology. They believe that using the word "archive" would suggest that it is a collection of primary sources when it is actually a body of newly created critical and historical scholarship.[68] Yet, they do acknowledge that the Orlando Project is, at the very least, an archive in the broader sense of a large collection of texts, in this case, about women writers. Interestingly, its creators also do not wish to describe their project as a database. Instead they call it a "textbase," declaring that the word "database" has "number crunching associations" that are foreign to their enterprise."[69] But whatever they choose to call it—archive, narrative, and database provide the language that allows them to conceptualize and describe their project.

Endeavors like Orlando are forcing the evolution of database into a more flexible tool for humanities scholars. The explicitly political goal of the Orlando Project to write women into British literary history was facilitated by the use of SGML encoding because it allowed its creators to specify and foreground those aspects of their texts that they wished to highlight and make searchable. Databases and encoded texts implicitly privilege some things over others and thus determine the kinds of questions one may pose. Unfortunately, in the digital environment, the nature of those limits is often invisible to the user. This is why many are troubled by Google's failure to reveal its ranking algorithms. Creating the tagsets and search engine for the Orlando Project and making the code explicit and visible was central to their project. Nevertheless, no one is declaring text encoding to be a panacea. As both the Orlando founders and others have argued, encoding has its limitations. However much one refines one's tagging, one is still forced to impose a level of specificity and explicitness on texts that, in the humanities at least, defy such clear cut-distinctions.

> **Databases and encoded texts implicitly privilege some things over others and thus determine the kinds of questions one may pose.**

Database and Bibliographic Control

The difficulty of imposing explicit structure on digital objects has also proven to be a challenge to the library community. For over a decade, librarians have been contending with new kinds of objects that defy standard forms of bibliographic description. The networked environment represents a shift toward the creation of information objects that are deeply embedded within their systems of hyperlinks and subject to recombination, relocation, and disappearance. This fluidity raises the question of the extent to which

libraries depend upon the existence of discrete, clearly bounded objects. This is not just an issue for librarians. Our methods for transmitting and preserving culture and for advancing and perpetuating knowledge presuppose the existence of discrete artifacts that can be described, organized, saved, and cited as separate and distinct entities.

In some ways, the instability of the bibliographic environment is not new. It has always been a challenge to distinguish among multiple versions, editions, translations, and adaptations of texts. But now the problem is greatly compounded by the possibility that any given digital document might be altered or erased. The item that one cites today may be a different or significantly revised object in a completely different location tomorrow. The greater instability of digital objects is a critical issue for the future of the scholarly record.

The irony is that, while computers are in some ways the ultimate memory tool, providing the ability to store and access immense bodies of information, they introduce new vulnerabilities. On the one hand, we face the problem of the persistence of digital data, such as potentially compromising personal information, e-mail, surveillance data, or sensitive financial or medical records. On the other hand, we face the problem of the potential loss of digital data and the vulnerability of the

> # Digital preservation might be an oxymoron. The digital both fosters and threatens the archival record.

digital record to manipulation, distortion, and erasure or loss through the supercession of hardware and software. Seamus Ross maintains that, after years of research into digital curation and preservation, we still face theoretical, methodological, and technological challenges to sustaining material held in digital libraries.[70] Digital preservation might be an oxymoron. The digital both fosters and threatens the archival record. The challenges of preservation are coupled with anxiety about precisely what digital information is being saved and by whom. We are uncertain about what should not be saved, what can and must be saved, and how to save it.

In today's digital environment, search engines are biased; library catalogs and database hyperlinks are sometimes inaccurate or unreliable. Nevertheless, more material is available to more people than ever. As Robert Darnton points out, the archive has long presented bibliographic challenges.[71] Scholars have traced concerns about information overload back to the sixteenth century. [72] The confidentiality of private information has long been an issue, though never on the scale fostered by today's huge databanks and social networking Web sites. Private ownership of knowledge is also not a new phenomenon. What we are seeing is not so much a rupture as a continuum. As the digital fosters the expansion of the archive, archival metaphor and archival theory proliferate. Through this process we implicitly conceive of the digital in terms of continuity. To theorize the digital through its archival effects is to acknowledge a range of forms that for centuries have served archival functions, providing a medium in which to create, store, and transmit culture and information. Although corporate ownership of intellectual property poses a threat to democratization of access, the movement from print to digital forms, like the movement from manuscript to print, has generally had a democratizing effect in terms of both content and access. Scholars surveying the technology of early printed texts also see continuity. A number of authors have explored the ways in which

searching, retrieval, and organizational techniques of modern computers have been anticipated and sought after for hundreds of years.[73] Investigating the ways in which Renaissance scholarship and finding aids anticipated modern computing, literary scholars describe the ways in which earlier forms such as commonplace books prefigure or anticipate the database.[74]

The archive has always been vulnerable. Libraries have been burned, bombed, and flooded. Public and governmental records have been lost or destroyed. Only a small volume of material survives from earlier periods such as classical Greece. But the digital archive is also at risk. It is threatened by the obsolescence of hardware and software and by the sheer mass of material in digital forms appearing and disappearing from networks on a daily basis. The notion of the archive is useful in theorizing the digital precisely because it carries within it both the ideal of preserving collective memory and the reality of its impossibility.

Marlene Manoff is senior collections strategist, MIT Libraries, Cambridge, MA; she may be contacted via e-mail at: mmanoff@MIT.EDU.

Notes

1. Marlene Manoff, "Theories of the Archive from Across the Disciplines," *portal: Libraries and the Academy* 4, 1 (2004): 9–25.
2. *Merriam-Webster's Collegiate Dictionary*, 11th ed., s.v. "Archive."
3. Ibid., s.v. "Database."
4. Manoff.
5. Edward Folsom, "Database as Genre: The Epic Transformation of Archives," *PMLA* 122, 5 (2007): 1571–9.
6. Lev Manovich, *Language of New Media* (Cambridge, MA: MIT Press, 2001), 218.
7. Jacques Derrida, *Archive Fever: A Freudian Impression*, trans. Eric Prenowitz (Chicago & London: University of Chicago Press, 1995).
8. Belinda Barnet, "Pack-rat or Amnesiac? Memory, the Archive and the Birth of the Internet," *Continuum: Journal of Media and Cultural Studies* 15, 2 (2001): 217–31.
9. Ibid., 229, 221.
10. Peter Stallybrass, "Against Thinking," *PMLA* 122, 5 (2007): 1581.
11. The Women Writers Project, Brown University, http://www.wwp.brown.edu/about/index.html (accessed July 15, 2010)
12. Ibid.
13. Matt Raymond, The Library and Twitter: An FAQ, Library of Congress Blog, posted April 28, 2010, http://blogs.loc.gov/loc/2010/04/the-library-and-twitter-an-faq/ (accessed July 16, 2010).
14. Randall Stross, "When History is Compiled 140 Characters at a Time," *New York Times*, May 2, 2010, http://www.nytimes.com/2010/05/02/business/02digi.html?scp=1&sq=when+history+is+compiled&st=nyt (accessed July 16, 2010).
15. Ibid.
16. Ibid.
17. Neil Rhodes and Jonathan Sawday, "Paperworlds: Imagining the Renaissance Computer: Knowledge in the First Age of Print," in *The Renaissance Computer: Knowledge in the First Age of Print*, ed. Neil Rhodes and Jonathan Sawday (London and NY: Routledge, 2000), 10.
18. Richard Lanham, "The Two Markets: Libraries in an Attention Economy," *Library Resources & Technical Services* 52, 2 (2008): 3–11.
19. Ibid.

20. Yehuda E. Kalay, "Introduction: Preserving Cultural Heritage through Digital Media," in *New Media and Cultural Heritage,* ed. Yehuda E. Kalay, Thomas Kuan, and Janice Affleck (London and NY: Routledge, 2008), 1–10.

21. World Heritage Forum, http://worldheritage-forum.net/en/2006/01/119 (accessed July 16, 2010).

22. Will Straw, "Embedded Memories," in *Residual Media*, ed. Charles R. Acland. (Minneapolis: University of Minnesota Press, 2007), 3–15.

23. Ibid.

24. Ibid.

25. Nick Thomas, "Tinted Love," *Boston Globe*, June 24, 2006, http://www.boston.com/ae/movies/articles/2006/06/24/tinted_love (accessed April 24, 2010).

26. Henry Jenkins, *Convergence Culture: Where Old and New Media Collide* (New York: New York University Press, 2006), 6.

27. Jerome McGann, "Our Textual History," *Times Literary Supplement*, November 20, 2009, 13.

28. Marlene Manoff, "The Materiality of Digital Collections: Theoretical and Historical Perspectives," *portal: Libraries and the Academy* 6, 3 (2006): 311–25.

29. N. Katherine Hayles, "Translating Media: Why We Should Rethink Textuality," *Yale Journal of Criticism* 16, 2 (2003): 263–90.

30. Roger Chartier, "Languages, Books, and Reading from the Printed Word to the Digital Text," special issue, *Critical Inquiry* 31, 1 (Autumn 2004): 147.

31. Laura Mandell, Preliminary Proposal for 18thConnect, 18thConnect, http://unixgen.muohio.edu/~poetess/NINES/ProposalPage.html (accessed July 16, 2010).

32. Stallybrass, 1580.

33. Ibid., 1581.

34. William Turkel, "The Search Comes First," Digital History Hacks (2005–2008): Methodology for the infinite archive, Weblog, Tuesday, January 1, 2008, http://digitalhistoryhacks.blogspot.com/2008_01_01_archive.html (accessed July 31, 2010).

35. Jonathan Freedman, "Whitman, Database, Information Culture," *PMLA* 122, 5 (2007): 1597.

36. Mandell.

37. *Journal of Library Administration* 47, 1/2 (2008), special issue, "Googlization of Libraries."

38. Norman Oder, "Library Groups Ask Justice Department to Supervise Institutional Pricing for Google Book Database," *Library Journal* (December 17, 2009), www.libraryjournal.com/article/CA6712243.html (accessed July 16, 2010).

39. Mandell.

40. Miguel Helft, "Wiki Citizens Taking on a New Area: Searching," quoting Wales, *New York Times* January 7, 2008, http://www.nytimes.com/2008/01/07/technology/07wiki.html (accessed July 16, 2010).

41. Jason Rody, "DFG/NEH: New Awards/ New Funding Opportunities," (7/9/09), National Endowment for the Humanities, Office of Digital Humanities, http://www.neh.gov/ODH/ODHUpdate/tabid/108/EntryId/111/DFG-NEH-New-Awards-New-Funding-Opportunities.aspx (accessed July 16, 2010).

42. JISC/NEH Transatlantic Digitization Collaboration Grants, National Endowment for the Humanities, http://www.neh.gov/grants/guidelines/JISC.html (accessed July 16, 2010).

43. i2010 European Digital Libraries Initiative: High Level Expert Group on Digital Libraries, Final Report on Public Private Partnerships (May 2008), http://ec.europa.eu/information_society/activities/digital_libraries/doc/hleg/reports/ppp/ppp_final.pdf (accessed July 16, 2010).

44. Ibid., 2.

45. Ibid., 28.

46. Alex Pang, End of Cyberspace, blog, http://www.endofcyberspace.com (accessed July 16, 2010).

47. Trebor Scholz and Paul Hartzog, "Toward a critique of the social web," Re-public: reimagining democracy, http://www.re-public.gr/en/?p=201 (accessed July 14, 2010).

48. Luke Treddinick, *Digital Information Culture: The Individual and Society in the Digital Age* (Oxford: Chandos, 2008), 165.
49. Manovich, 219.
50. Folsom, 1571.
51. Ibid., 1575–6.
52. Ibid., 1572.
53. Ibid., 1574.
54. Freedman, 1599.
55. Ibid., 1599–600.
56. Folsom, 1575.
57. Jerome McGann, "Database, Interface and Archival Fever," *PMLA* 122, 5 (2007): 1588.
58. Ibid.
59. Manovich, 234–6.
60. N. Katherine Hayles, "Narrative and Database: Natural Symbionts," *PMLA* 122, 5 (2007): 1603-1607.
61. Ibid.
62. University of Alberta, Orlando Project: A History of Women's Writing in the British Isles, University of Alberta, http://www.ualberta.ca/ORLANDO/ (accessed July 16, 2010).
63. Susan Brown, Patricia Clements, Isobel Grundy. "Sorting Things In: Feminist Knowledge Representation and Changing Modes of Scholarly Production," *Women's Studies International Forum* 29, 3 (2006): 317–25.
64. Ibid., 319.
65. Caroline Levander. "The Changing Landscape of American Studies in a Global Era," in *Working Together or Apart: Promoting the Next Generation of Digital Scholarship*, *Report of a Workshop Cosponsored by the Council on Library and Information Resources and National Endowment for the Humanities* (Washington, D.C.: Council on Library and Information Resources, 2009), 27–34.
66. Ibid., 29.
67. Ibid., 31.
68. Susan Brown et al., "An Introduction to the Orlando Project," *Tulsa Studies in Women's Literature* 26, 1 (2007): 127.
69. Ibid.
70. Seamus Ross, *Digital Preservation, Archival Science and Methodological Foundations for Digital Libraries* (keynote address, 11th European Conference on Digital Libraries [ECDL], Budapest, September 17, 2007), http://www.ecdl2007.org/Keynote_ECDL2007_SROSS.pdf (accessed July 16, 2010).
71. Robert Darnton, "The Library in the New Age," *New York Review of Books* 55, 10, (June 12, 2008), http://www.nybooks.com/articles/21514 (accessed July 16, 2010).
72. Daniel, Rosenberg, "Early Information Overload," *Journal of the History of Ideas* 64, 1 (2003): 1–9.
73. Rhodes and Sawday.
74. Ibid.