The Digital Record and the Future of Libraries

by

Marlene Manoff, April 2009

Academic and research libraries support scholarship, history and memory. They enable us to define who we were and thus who we are. But for the past two decades, the digital has been altering and remaking library cultures and organizations. Libraries are being transformed as they absorb and assimilate new kinds of cultural objects and records. The digital is proliferating while the publication and importance of print persists; analog and digital cannot be viewed as simple dichotomies in need of resolution in some pure digital future. Libraries cannot avoid the messiness and ambiguities of the current moment.

All scholarship is implicitly a negotiation with, an interpretation of and a contribution to the historical record. While libraries bear a responsibility to both the present and the future to preserve that record, they have less control than ever over the increasingly expensive, unstable and restricted electronic information products that they provide. I've written elsewhere about the implications of selection decisions made by libraries, archives and museums.[1] In the current environment, however, the question is not simply what libraries choose to acquire; increasingly it is about whether the material they acquire today will continue to be available twenty or a hundred years from now. If no action is taken to preserve digital material, the legacy of today's academic and research libraries may well be a vast hole in the intellectual record.

Libraries are being tested on a variety of fronts. Digital objects are unstable and resist the kind of bibliographic control libraries have traditionally provided. Despite years of work, we are not close to solving the problems of assuring digital longevity. The challenges are technical and economic as well as social and organizational.[2] Moreover, funding has not kept pace with the

explosive demand for access to new digital products, many of them so expensive as to dwarf library budgets. A current library subscription to, for example, Compendex or INSPEC, two major indexes covering different aspects of engineering, may cost over $40,000 each per year. Libraries typically subscribe to scores of databases covering a wide variety of professional and academic fields. The current economic crisis has only exacerbated the expense and difficulty of providing these digital resources.

Access and preservation, two key historical functions of academic and research libraries, are more difficult to reconcile in a digital environment. Digitization provides increased access to both contemporary and historical material, but the vulnerability of the digital record poses significant challenges to the ability of libraries to insure the integrity and sustainability of that record. With pressure to acquire as much digital material as possible, libraries often choose to outsource metadata creation, to lower their standards for bibliographic records and to put off consideration of or downplay questions of digital longevity. Even the Library of Congress wants to decrease its role as one of the largest creators and distributors of bibliographic records in the world in order to devote more resources to supporting digitization of unique holdings.[3] Libraries are experiencing pressure to relinquish their role in safeguarding the historical record in part because they believe they must put their limited resources into purchasing, creating or enhancing access to digital objects.

As electronic journals and library databases have come to dominate library budgets and holdings, it has become impossible for libraries to keep accurate records of what they are providing their users. Researchers have found that many vendors of journal database packages provide selective coverage of the material they ostensibly include in full-text. Ronald Banks surveyed the literature on journal database coverage and comparison and his findings are not

reassuring.[4] Banks cites a study that found that ProQuest and EBSCO, two key library vendors, provide different numbers of articles from the same journals. He also cites the work of two researchers who compared print versions of journals to the versions provided in full-text databases. They found that coverage of journals in database packages is subject to change without notice, that there is often significant time delay in the inclusion of new issues, and that only 62% to 89% of the content of print journals is included. Database vendors do not guarantee that their products will remain consistent over time. Items disappear from library databases without warning, material may be added or subtracted and libraries may thus be providing access to a different collection of digital objects from one day to the next.

This state of affairs has wreaked havoc on conceptions of bibliographic control that have been central to librarianship in the modern period. The function of bibliographic control is to insure that every item has its unique place in an organizational arrangement that allows for systematic searching, discovery and retrieval. But the web plays havoc with the notions of control, order, fixity and hierarchy that are at the heart of the bibliographic enterprise. In the print environment on any given day, between card catalogs, journal receipt records and circulation records, libraries could determine pretty accurately what materials were available to their users. This is not the case in the electronic environment. Libraries are not able to maintain records of each issue of each new electronic journal they "receive" or every item added to the databases they provide to users. Many libraries now subscribe to thousands of electronic journals and merely take it on faith that each new issue will appear at the appropriate time and that past issues are still available. There is no way for them to ascertain whether all the items for which they have paid remain accessible. And even if libraries could verify the appearance of each new issue of every journal to which they subscribed, there would still be no guarantee that in an hour,

a week or a month, all the issues of all the journals would remain available. The digital environment has vastly decreased libraries' ability to maintain an accurate inventory of their resources.

The complexity of the problem is suggested by the statement in the new 2009 7th edition of the *MLA Handbook for Writers of Research Papers*, that the Modern Language Association (MLA) no longer recognizes print as the default medium. This would seem to suggest an increased confidence in the digital record. But at the same time that the MLA has ceased to privilege print versions, it expresses scant faith in the ability to cite digital versions. The new handbook does not recommend the inclusion of URLs when citing web-based resources because they have proven "to have limited value…for they often change, can be specific to a subscriber or a session of use, and can be so long and complex that typing them into a browser is cumbersome and prone to transcription errors."[5] And if, as the handbook suggests instead, readers are best off typing authors and titles into search engines to locate cited works, chances are increased that they will find different or multiple versions or none at all.

The instability of the digital record has profound implication for scholarship. Knowledge in every discipline is built on a historic foundation. Each generation of scholars builds on the work of its predecessors. This requires the maintenance of a stable and reliable system of citation that enables researchers to trace ideas and discoveries back to their original sources. Scholars must be able to document the evolution of the research and theory that underpins their own work. But anyone consulting electronic citations in or to electronic journal articles, conference proceedings or other networked information soon realizes that a significant portion of these links are broken and no longer point to the original source. For decades libraries have played a central role in maintaining the accuracy and precision of bibliographic records. But traditional systems

of documentation have yet to be successfully transferred to the digital environment. In a study of what the authors call "the half-life" of internet references cited in communication journals, they analyzed 1126 URLs in articles that appeared between 2000 and 2003. They found that 39% no longer remained accessible.[6] Michael Bugeja, one of the authors of this study, was prompted to undertake it and several others after it was found that in a final manuscript of a book he was writing for Oxford University Press, 30% of the web-based footnotes no longer functioned forcing him to rewrite major portions of the book.[7] If a third or even a quarter of the Web-based material cited in academic and scholarly papers disappears or is no longer accessible because of link rot, we are confronting a significant threat to research as we have known it. Embedded citations and hypertext links constitute one of the major advantages of networked scholarship, but it is precisely the decay of those links that threatens the scholarly enterprise. These contradictions demand serious attention. Some suggest that bibliographic control can be reasserted over the flux that is the networked environment through the development of more advanced technologies; others think the loss is more than compensated for by the benefits and that we will eventually adjust to this new world.

Still, there is a growing recognition, even among mainstream writers that we should be concerned that the digital world is less reliable than the print one. Nicholas Carr, for example, recently used the example of Amazon's Kindle to illustrate the troubling fact that digital text distributed through a network provides no guarantee of authenticity.[8] As Carr explains, Amazon automatically updates the Kindle and can delete or edit the text of a book one has already bought and read leaving no trace of the original version. You may think your Kindle contains the book you just read but it well may have been invisibly altered. Electronic books, like electronic

journals and databases, can be difficult to authenticate and are vulnerable to unmonitored modifications and deletions. Printed books are not susceptible to these kinds of manipulations.

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. The digital record is particularly susceptible to distortion and erasure. The digital both fosters and threatens the historical record. The loss of stability is coupled with an ever-growing expansion of digital materials. A further irony is that digitization allows us to recover the past but in doing so alters and reframes it. Artifacts from different eras and geographic locations are brought into a largely undifferentiated present. Computer memory expands while the present appears to shrink through accelerating cycles of innovation and obsolescence. We have an insatiable appetite for cultural objects that connect us to a shared past. Enterprises like Google Books, the American Memory Project, online museum collections and digital heritage sites as well as the huge variety of older film, music and television programs available in digital formats allow us to experience a connection to other times, people and places. Will Straw describes a vast market for digitized historical artifacts like music, film or texts packaged as themed and annotated collections. These collections not only make the past more visible but they create new cultural canons and new readings of history.[9] Digital technology creates an appetite as well as a market for the historical objects it delivers and recontextualizes.

But while more historical material is available to more people than ever, the long term or even short term viability of such material remains in doubt. In the case of libraries, part of the problem is that much of the digital material they offer does not reside on servers they control. Users can access material because libraries either purchase access or purchase material outright. But even library ownership does not guarantee long term access, especially since much of this

material sits on the servers of private companies. Who knows whether ProQuest, Cengage Learning, LexisNexis, Google or any of the companies that sell or lease digital products to libraries will be around in 25 years? Online library catalogs (OPACS) are rife with links to networked resources that are subject to decay.

Libraries are counting on two nonprofit enterprises, LOCKSS and Portico to preserve electronic journals and possibly additional content. LOCKSS and Portico may provide long-term access to the thousands of journal volumes they are currently ingesting. But only after most academic and research libraries have cancelled thousands of print subscriptions has the library community undertaken in-depth assessments of these repositories. The Center for Research Libraries (CRL), a consortium of North American universities, colleges and independent research libraries recently announced that in 2009 it will be assessing Portico as part of a project "to promote understanding of and, where justified, confidence in, digital repositories."[10] CRL is also assessing the capabilities of LOCKSS and will present its findings in November 2009. At this point, no one knows how much contemporary material will be successfully migrated and available in fifty years. Libraries cancelling print versions of their journals are subjecting themselves and the historical record to new risks. This is a conscious choice in an environment that provides few attractive options.

Even the largest and most prestigious libraries have embarked upon the elimination of print subscriptions when digital versions are available. The Association of Research Libraries (ARL), an organization of 123 of the largest research libraries in the U.S. and Canada, issued a statement on February 19, 2009 acknowledging that research libraries have substantially reduced their print subscriptions.[11] The document also claims that given the current economic situation, "most research libraries would welcome the complete elimination of print publishing…if

acceptable preservation strategies are in place, such as archiving with third parties." But again, long-term digital archiving solutions remain elusive. So while the ARL document does acknowledge the responsibility of libraries in the stewardship of the scholarly record and the current risk of "the loss of important scholarly content," its priorities are elsewhere. Overall, research libraries have chosen to prioritize the dissemination of the electronic record over the preservation of either the print or electronic one.

Most research libraries are committing fewer resources to supporting their print collections so they can increase support for digital resources. This makes a certain amount of sense given the growth of electronic resources, but it does mean fewer dollars to buy books and fewer dollars to catalog and process print material. Yet books remain a key mode of presenting and disseminating research and scholarship in many disciplines. In a period of enforced economic retrenchment, it is seductive to believe that print is dying. But in some cases the economic crisis has become an excuse for moving quickly to curtail support for print resources. There is a considerable body of library panic literature proclaiming the end of libraries should they fail to quickly transform themselves into fully contemporary digital enterprises. Many who claim they are bowing to the inevitable may well be precipitating what they claim they cannot alter. If, for example, libraries preemptively send all their books and journals to storage, they are essentially training users to manage the change that libraries claim is inevitable. Staff and patrons who continue to insist upon the importance of being able to browse and consult onsite print collections are often made to feel as if they are somehow impeding a better future. Decisions about the appropriate balance of print and digital resources must be made in a deliberate way with full acknowledgement of the long-term implications and not just the short term savings.

Many individual librarians have sought to provide counter arguments or to complicate this media determinism. Thomas Mann[12] at the Library of Congress and John Buschman[13] at Georgetown University have both written at some length about the perils of so simple a view of the current environment. But the library literature in general tends to be extremely practical or technical, offering empirical studies of tools and practices, encouragement to adopt the latest web 2.0 applications, and many exhortations to get with new media or die with old media. The current moment of media change is messy, multi-faceted and ambiguous. Simple narratives of media determinism are inadequate. The future will be shaped by a variety of forces including the choices made by academic and research libraries which cannot pretend to be neutral. Over the past two decades, by investing the bulk of their resources in expanding coverage of journals in science, technology and medicine (STM) that have grown exponentially in price, libraries have consistently shrunk the portion of their budgets devoted to monographs. Libraries may claim that they had no choice in this matter, but the reality is more complicated. Library decisions undoubtedly played a role in creating the current situation in which the library market for monographs has contracted. This in turn has had a tremendous impact on humanities and social science scholarship where there are many fewer options for publishing print monographs. Only now that budgets are tighter than ever are libraries finally mounting serious resistance to predatory STM pricing.

In a few key areas libraries have recognized that they can in fact intervene to shape the future. Librarians are taking progressive stands on intellectual property rights, privacy issues and net neutrality. They are also on the cutting edge of advocacy work, educating scholars and politicians about the stakes of current copyright practices and legislation and lobbying for broad access to federally funded research as well as government information. Many libraries are

working to help faculty reclaim control of their intellectual output. Locally, Harvard, MIT, Tufts and B.U. are all making strides in this direction. But the complexity of the current moment, the proliferation of the digital, the persistence of the analog and the bleak economic environment threaten libraries' abilities to devise solutions or to overcome obstacles to establishing a more multifaceted response to the digital analog mix and the competing demands of access and preservation.

Further complicating the issue for academic and research libraries is the changing nature of scholarship and the way data is increasingly embedded in that scholarship. Especially in the sciences, but also increasingly in the humanities and social sciences, researchers are amassing large bodies of data, using technology to perform a variety of analyses and simulations. One example outside the sciences is the creation of digital reconstructions, visualizations and models of cultural heritage sites. New digital affordances enable amazingly detailed and multi-dimensional recreations of ancient objects, structures, and even entire cities. The largest projects, however, are in the sciences where the challenges are greatest in terms of curating and preserving massive amounts of data.

Libraries are beginning to appreciate the importance and challenge of these developments. They are recognizing the need for more collaboration in order to find ways to store and maintain huge collections of data as well as the software to manipulate it. Libraries are seeing data curation as a new role they may take on to maintain their centrality to the scholarly enterprise. In this context, libraries are acknowledging their responsibilities to steward the intellectual record.[14] Yet on a library by library basis, there is a growing readiness to outsource and reduce metadata creation whether for print or digital objects. In the words of Martha M. Yee, a cataloger at U.C.L.A., "it appears that the ALA [American Library Association] is now

dominated by library administrators with shrinking budgets who know very little about the complexities of bibliographic control (other than its expense) and who wonder if the fact that undergraduates are in love with Google might not provide an excuse for libraries to dispense with the information-organization part of their budget entirely."[15] Libraries can't have it both ways. If they compromise their standards on metadata creation and don't invest in long-term digital preservation, they will not be in a position to take responsibility for data curation.

Part of the difficulty we confront in this extremely complex environment may have to do with the fact that libraries are practical enterprises and the field of librarianship has not developed the kind of abstract disciplinary foundation on which to build more compelling and multifaceted analyses and strategies. As Sydney Pierce has pointed out, there are no "dead Germans" to provide theoretical and philosophical underpinnings for library science.[16] In other disciplines with stronger theoretical foundations, writers have produced more nuanced readings of the current moment, recognizing, for example, that digital and analog are mutually redefining. Martin Hand is one of many media theorists addressing the interconnection of digital and analog culture and the danger of theories of the digital as either revolutionary change or illusive fantasy. He goes so far as to question the notion of media transition as currently understood as it "tends to imply the loss or replacement of analogue (sic) by digital."[17]

But analog is not going away. As Charles Acland points out, the danger of focusing solely on media change is that it comes "at the expense of taking account of continuity, fixity, and dialectical relations with existing practices, systems, and artifacts."[18] We live in a world where the virtual coexists with the real. Even Clifford Lynch, a leading voice in the field of library technology, predicts that "traditional published materials will continue to be important," that "[f]or research libraries, change is going to be somewhat gradual," and that "the proportion

of the overall collections in digital form will grow fairly slowly."[19] Libraries need to carefully parse current trends. They must balance the desire to stay relevant, to innovate and provide new products and services, to offer the coolest websites and the niftiest new library applications, with their responsibilities to current and future generations of scholars. The demands of the digital environment should not be an excuse to compromise on standards and to dismantle traditions of legitimating knowledge thorough the construction of the bibliographic record. New technologies may one day provide better or cheaper ways to accomplish this work, but an accurate bibliographic record cannot simply be dismissed as an artifact of some superseded print paradigm.

Library leaders believe they must embrace risk in order to support digital resources. This supposition goes to the heart of the dilemma. Almost a decade ago Nicholson Baker raised the question of the library communities' responsibility to steward the historical record.[20] But library leaders believe that they must innovate in order to support new kinds of scholarship and new publishing arrangements. And they are choosing to look forward rather than back. As noted earlier, even the Library of Congress, the largest research library in the country, has made it a top priority to finance digitization of unique collections by cutting support for print acquisitions and cataloging. Libraries need a new calculus to balance competing demands. Supporting and broadening access to digital material should not come solely at the expense of diminished bibliographic control, diminished access to print and diminished preservation of the historical record whether in analog or digital form.

---

[1] Manoff, Marlene. "Theories of the Archive from Across the Disciplines." *Portal: Libraries and the Academy* 4:1 (2004): 9-25. (http://dspace.mit.edu/bitstream/handle/1721.1/35687/4.1manoff.pdf?sequence=1)
[2] Ross, Seamus. *Digital Preservation, Archival Science and Methodological Foundations for Digital Libraries*, Keynote Address at the 11th European Conference on Digital Libraries (ECDL), Budapest (17 September 2007).

[3] "On the Record: Report of the Library of Congress Working Group on the Future of Bibliographic Control." January 9, 2008 (http:www.loc.gov/bibliographic-future/news/lcwg-onthe**record**-jan08-final.pdf).
[4] Banks, Ronald A. "Decision-Making Factors Related to Bibliographic Database Cancellation." *Behavioral & Social Sciences Librarian* 25.1 (2006): 93-109.
[5] Redden, Elizabeth. "Style: MLA, Updated." *Inside Higher Education*, March 11, 2009 (http://www.insidehighered.com/news/2009/03/11/mla).
[6] Dimitrova, Daniela V. and Bugeja, Michael. "The Half-life of Internet References Cited in Communication Journals." *New Media & Society* 9.5 (2007): 811-826.
[7] Bugeja, Michael. "The Five 'W's' and 'H' of the Half-Life Phenomenon." (http://www.halfnotes.org/our_history.html).
[8] Carr, Nicholas. "The Automatically Updatable Book." Rough Type: Nicholas Carr's Blog, February 11, 2009 (http://www.roughtype.com/archives/2009/02/the_automatical_1.php).
[9] Straw, Will. "Embedded Memories." In *Residual Media*, edited by Charles Acland. Minneapolis: University of Minnesota Press, 2007.
[10] "2009 Certification and Assessment of Portico and HathiTrust." http://www.crl.edu/content.asp?l1=13&l2=58&l3=181
[11] Association of Research Libraries. *ARL Statement to Scholarly Publishers on the Global Economic Crisis.* February 19, 2009 (http://www.arl.org/bm~doc/economic-statement-2009.pdf).
[12] Mann, Thomas. "What is Going on at the Library of Congress?" Prepared for AFSCME 2910 The Library of Congress Professional Guild. June 19, 2006 (http://www.guild2910.org/AFSCMEWhatIsGoingOn.pdf), "Peloponnesian War and the Future of Reference, Cataloging, and Scholarship in Research Libraries." Prepared for AFSCME 2910 The Library of Congress Professional Guild. June 13, 2007 (http://guild2910.org/Pelopponesian%20War%20June%2013%202007.pdf), and "'On the Record' but Off the Track: A Review of the Report of the Library of Congress Working Group on the Future of Bibliographic Control, with a Further Examination of Library of Congress Cataloging Tendencies," March 14, 2008. (http://www.guild2910.org/WorkingGrpResponse2008.pdf
[13] Buschman, John E. "Introduction: Information Technologies and Libraries – Why Do We Need New Critical Approaches?" and "Conclusion: Just How Critical Should Librarianship Be of Technology?" both in *Information Technology in Librarianship: New Technical Approaches*, edited by John E. Buschman and Gloria J. Leckie. Westport, Connecticut: Libraries Unlimited, 2009 and Buschman, John E. *Dismantling the Public Sphere: Situating and Sustaining Librarianship in the Age of the New Public Philosophy*. Westport, Conn.: Libraries Unlimited, 2003.
[14] Duderstadt, James J. "Possible Futures for the Research Library in the 21st Century." *Journal of Library Administration 4:3* (2009): 217-225.
[15] Yee, Martha M. "'Wholly Visionary': The American Library Association, the Library of Congress, and the Card Distribution Program." *LRTS* 53:2 (April 2009): p. 75.
[16] Pierce, Sydney J. "Dead Germans and the Theory of Librarianship." *American Libraries* 23.8 *(*September 1992): 641-643.
[17] Hand, Martin. *Making Digital Cultures: Access, Interactivity, and Authenticity*. Aldershot, England: Ashgate, 2008, p. 165.
[18] Acland, Charles R. "Introduction." *Residual Media*, edited by Charles Acland. Minneapolis: University of Minnesota Press, 2007.
[19] Lynch, Clifford and Carleton, Don E. "Lecture: Impact of Digital Scholarship on Research Libraries." *Journal of Library Administration* 49:3 (2009): 227-244.
[20] Baker, Nicholson. *Double Fold: Libraries and the Assault on Paper.* New York: Random House, 2001.