

# Revolutionary by Design: HathiTrust, Digital Learning and the Future of Information Provision

*Kent LaCombe*

The advent of information digitization led inexorably to dramatic changes in the methods of information production, dissemination, access, interpretation and application. For academic libraries, the tried and true models of print acquisition, ownership and delivery share limited procurement capitals alongside a burgeoning landscape of digital data resources. Networked data is now accessible and searchable at quantities and speeds never possible in a print dependent environment. However, the shared world of print and electronic data continues to present considerable challenges that will likely remain unresolved for the foreseeable future. Libraries exist in a reality of duplicate acquisitions across formats, competing commercial vendor interfaces, limited purchasing resources and evolving patron demands. Academic libraries must resist the urge to mirror the competitive commercial entities upon whom they increasingly rely. Libraries must support one another and embrace an inclusive vision of human education, within and beyond campus borders and like-minded academic consortia, toward a holistic, shared vision of intellectual enrichment supportive of open access and valuing library owned and library administrated resources.

An early and critical component of the digital information landscape was JSTOR, a revolutionary, non-profit resource that changed the face of print journal acquisition and preservation while it concurrently transformed the way faculty and students constructed their literature reviews. First launched in 1995 through pioneering efforts by the University of

Michigan, the digital archive quickly became an essential tool for countless researchers.<sup>1</sup> Today, very few scholars make it through an entire program of study without consulting JSTOR. HathiTrust is poised to become a resource of even greater global accessibility, research import and recognition. The collaborative, massive and growing digital repository is forever changing the way academic libraries acquire, preserve, share and provide digital information.

The accessibility of digital information resources affected all fields of research in dramatic and sometimes different fashions over time. Occasionally we all hear fellow academics refer to certain fields as slow to adopt new technologies or various disciplines as print dependent, while others supposedly no longer use print. Broad brush assumptions about the mechanics of any field of interdisciplinary research are seldom correct, while at the same time the critical importance of timely access to digital data is impossible to dismiss for any area of serious inquiry. Numerous individuals in what are still occasionally (and often unfairly) labeled print dependent fields long recognized the value of digital formats and access. While working as an undergraduate humanities scholar in the 1990s, select faculty encouraged me to explore what were then cutting edge electronic resources in history. Those scholars were keenly aware of the revolutionary nature of products such as ABC-Clipo's CD-ROM versions of *America: History and Life*.<sup>2</sup> Early adopters of those innovative resources helped ensure their students were literate in the emerging world of digital information procurement and use.

---

*Kent LaCombe is Assistant Professor of Libraries, University of Nebraska, Lincoln, e-mail: klacombe2@unl.edu*

At the same time, there were many skeptics. Critics of digital resources were widespread. In his 2002 analysis of a JSTOR surveys, Kevin M. Guthrie pointed out that 74% of humanities faculty opposed replacing print journals with digital archives.<sup>3</sup> In JSTOR's early years, more than a few students were discouraged from using its resources. While discomfort with technology sometimes generated such skepticism, resource longevity and reliability were legitimate, major concerns at the end of the last century, and they remain so today. While JSTOR quickly evolved into arguably one of the most recognized, user friendly and reliable resources for mining vast amounts of scholarly information, other data interfaces were unintuitive or became design-moribund, changed ownership or juggled content. Today, data stored and shared from university repositories and various commercial subscription portals competes with a readily available and familiar one-size-fits all Google interface that scores of students and more than a few faculty turn to when researching various topics. Thus, the ease of navigation through a university's portal is as important today as the digital resources it leads to. In the muddy and competitive world of for-profit and nonprofit providers, full text, citations, abstracts, varied federated searches, open source, peer review, institutional digitization, digital rights management and dozens of other factors, there is no easy roadmap toward effective information procurement, access and stewardship. Nor can one simply adhere to what worked in the past nor alternately throw all their resources toward an assumed, unknown future. What if we'd all discarded print for CD-ROMs? On the other hand, there is no such thing as a print dependent research field any more than there is, currently at least, entirely digitally dependent endeavors. Research is information dependent. Increasingly much of that information is contained in digital formats, while additional vast percentages remain in print. We must continue to serve the needs of today as much as we prepare for and embrace the changes of tomorrow. How is that best accomplished? How do we navigate the confused, choppy waters of information provision as we rapidly

transition across a changing format landscape with limited procurement resources?

One avenue of acquisition libraries are embracing is ebooks access and purchasing. Digital books are steadily gaining ground in many collections. Unfortunately, at all levels of acquisition and access standardization is absent. Different publishers follow different timetables for releasing digital versions of their publications. Different ebook vendors utilize individual and competing interfaces in order to access the resources they control. They embed restrictive digital rights management coding to protect both copyright and corporate financial interests. Libraries pay for individual ebook titles. Yet even after acquiring titles libraries often pay additional access fees to individual vendors and must utilize one of the numerous corresponding ebook access portals, replete with various limitations on how items can be viewed, downloaded, printed, marked up or otherwise utilized.<sup>4</sup>

Institutions of every stripe are often slow to adopt or adjust to changes in established practices. Libraries are no different. Changing established practices can be especially challenging given past success and firmly established values associated with traditional service models and collections. Often stakeholders seize the ramparts on distinct "sides" of the service question and a false dichotomy is created. In his essay *The Myth of Browsing*, Donald A. Barclay considered the vociferous response of the allies of tradition and printed stacks at Syracuse University. Barclay argued that defenders of printed journals and monographs typically found their arguments on two pillars. The first is a defense of serendipitous discovery in the stacks through physical browsing of collections. The second was a perceived, inherent value- what Barclay termed "a vibe, an ambiance, a holiness" to being surrounded by printed works. He argued that the heavily used items would be absent from the stacks, so serendipitous discovery would be like "hitting the sale tables on day three of a three day sale."<sup>5</sup> Barclay went on to describe the daunting mechanics of physical library cataloging and shelving. He illustrated twentieth century models that worked well given the physical and

technological infrastructures of the time—but times have changed.

When reading Barclay's critique one sees a clear vision supportive of our digital future. However, the demands of print collection defenders, while perhaps aggravating to some digital proponents, are often grounded in their experience and shared reality and more importantly the shared realities of many of our patrons. Barclay too conceded we must retain a "prime spot" for the printed book, though like the rest of us, he grappled with what that means.<sup>6</sup> Libraries did not archive enormous print collections to be inefficient. Having done research in both the humanities and the sciences, I can attest to serendipitous discovery on a personal level while shelf-browsing a number of fields, although it is admittedly, in my opinion, neither an efficient nor time-saving method of inquiry.

On the other hand, I will not argue the merits of collection ambiance for ambiance's sake. We each instill our personal value to the objects we care about. As an undergraduate student of history I came in to my chosen program of study heavily influenced by social upbringing and popular culture. While completing one of my first research trips to a regional archive I was given the opportunity to hold an elementary textbook cover once used by the future General George Armstrong Custer. At the time that object held tangible and intangible meanings for me that were rooted both in a woefully incomplete understanding of my field as well as acculturated biases. That same object would have meant something entirely different to someone else and to many it would mean nothing at all. Which emotional or intellectual response is correct? Occasionally we may hear colleagues speak to the unique experience associated with holding physical artifacts, including books, in one's hands. How are digital works less valid? How is the value countless scholars place on a digital work somehow less valid, less tangible, or less meaningful?

Personal and educational experiences and the values we place on them change, yet they are all critical components of intellectual growth. The complex yet flawed conclusions about national history I held as a

youth are recalled when I think of holding that physical book cover. My current and constantly evolving understanding of history bears little resemblance to the years of my undergraduate education. Holding that same object would mean something entirely different to me today.

Digital provides the same opportunities for discovery and reflection, but does it faster and on a larger and far more readily accessible scale. Two years ago while demonstrating HathiTrust to a group of undergraduate students studying medical history; I literally stumbled across a scanned text that included a handwritten dedication to Henry Ford. It was serendipitous discovery—digital style. The text was a 1927 edition of *Scientific Fasting; The Ancient and Modern Key to Health* by Linda Burfield Hazzard.<sup>7</sup> The class knew who Henry Ford was. Linda Hazzard was a mystery to them. Hazzard was a medical quack and serial killer who starved patients to death at her sanitarium in the state of Washington. That chance discovery provided an opportunity to give a memorable demonstration of an incredible digital repository and display a unique resource none of us were likely to ever view without something akin to HathiTrust. I will never forget the experience, and I believe many of the students will remember it as well. Digital objects are as valuable as we perceive them. They are the artifacts of the future, and can preserve forever what might otherwise be lost to the passage of time.

Despite that, the ramparts for a mounting a complete digital defense remain unfinished. We have not reached a stage of sufficient digital infrastructure where academic libraries can comfortably discard or relocate print to such an extent that day-to-day research is hampered. To do so in the present day endangers the primary role of an institution of higher learning. The digital age is creating a generation that, for better or worse, expects instant access. It is with some irony that current expectations built using digital formats can concurrently fuel a need for print access. In technologically developed societies human populations are being conditioned, for better or ill, to expect instant access. If key materials are intention-

ally or accidentally relocated countless scholars will not bother to try accessing them. The hurdle of off-site storage is made doubly tall when additional time is added for delivery. Low circulation materials will quickly become zero circulation items when they are located in off-site storage. Print becomes archived and mysterious. It is, ironically, a return to a kind of closed stacks print library similar to those critiqued in Barclay's article. On the other hand, discarding print collections is an even more problematic affair, especially if digital replacements are not available and substantial investments in digital formats and resource providers is not possible. Librarians are not equipped to discern all of the potentially important texts across large departments of served faculty at major research universities. Much like library science, all major fields of research are enormous and have countless studies important to their purveyors. We are at a critical turning point in library services when dramatic change must be balanced by decisive but thoughtful and informed decision making. Sometimes even the best intentioned, largest plans are abandoned or reformulated at the last minute. Just last May the New York Public Library halted a proposed \$150 million dollar renovation in the face of mounting criticism from, among others, professional researchers who use the library's significant, on-site, non-circulating print collections. Many of those collections were slated for relocation.<sup>8</sup>

At most universities, space is also a contested premium. When library space is recommissioned, it is incumbent upon our current generation of library administrators to be cognizant of the library's changed and changing role while remaining informed, firm educators and champions of their profession in and outside their campus communities. Librarians aren't the only ones observant of our professional evolution. We must educate others of our changing profession and its continuing needs, lest administrators representing other campus issues look to "wasted" library spaces for easy answers to their own issues. Academic libraries are grappling with a host of questions about the use of space and how to best serve the diverse

needs of their constituents. Inaction is not an option. What opportunities offer possible solutions?

In our age of limited resources coupled with growing institutional pressure toward enrollment and grant dollar competition, academic libraries must askew competition and instead become cooperative ventures. Consider now, HathiTrust. The encouraging model and ultimate possibilities embodied by the still relatively young cooperative provide an excellent opportunity for both our collective support and eventually, emulation. HathiTrust was launched in 2008. It developed from cooperative roots born of the Committee on Institutional Cooperation (consisting of the Big Ten universities plus the University of Chicago), the University of California Library System and Google.<sup>9</sup> The University of Virginia joined the following year. HathiTrust embodies a present-day example of successful, increasingly global, library cooperation. Many of the partnership's institutional members are actively scanning their materials for inclusion in the shared archive. Documents and monographs that are out of copyright can be accessed via HathiTrust's highly customizable digital interface, by anyone, anywhere in the world with the technology infrastructure to do so. Mass scanning of collections for consideration and inclusion in the archive is one way libraries can ensure their print collections are preserved. Nor should these activities be conceptualized as print limited. The HathiTrust consortium is investigating the incorporation of additional formats in to the digital archive.<sup>10</sup>

Low print availability, unique items, out-of-copyright collection materials and other institution specific resources present a plethora of avenues for beginning investigation of in-house digital preservation. Mass scanning can provide redundant backups through digital duplication. Physical items can then be moved to on-site or shared storage with confidence and in some cases deaccessioned. In such a process, no collections, including potentially unique or under-represented items, are accidentally and permanently lost. As we are in a transitional period in formats and services in academic libraries, increased digital collection creation is an important avenue ripe for

investment. The 2014 U.S. Second Circuit Court of Appeals decision in *Authors Guild, Inc. v. HathiTrust* upholding fair use was a timely victory for cooperative academic information access, preservation and use.<sup>11</sup> In an age when education provision is too often hampered and influenced by economic variables and copyright question marks at all levels, the material incentives of increased cooperation and mutual support should be obvious. A cooperative agenda can be applied to other areas of information acquisition and delivery, such as resource acquisition. In addition, it is critical that academic institutions present a united front when confronted with direct challenges to their educational mission and not stand on the sidelines as similarly interested colleagues fight our battles over copyright, fair use, licensing, access and myriad of other issues that are and will continue to arise as the face of higher education evolves.

Over the last few years, HathiTrust developed into a resource repository as revolutionary for the future of information provision as JSTOR was when it first appeared late in the last century. HathiTrust, like JSTOR, is a nonprofit. It is developed, owned and operated by a growing consortium of educational partners. At this writing, its rapidly growing holdings number over thirteen million volumes, with approximately five million in the public domain. In the words of HathiTrust Assistant Director Jeremy York, it is “one of the largest research library collections in the world.”<sup>12</sup> Additionally, a growing percentage of public-domain monographs and other publications can be downloaded as standalone full-text PDFs.<sup>13</sup>

HathiTrust’s content grew exponentially in a relatively short period of time. As the number of consortium members grows, the percentage of those members who choose to actively embrace, participate defend in the opportunity to provide for the needs of their affiliated campus stakeholders and the greater public good will grow with it.<sup>14</sup> The HathiTrust consortium is placing resource creation, management, ownership and preservation back in the hands of the educational community on a global scale. HathiTrust is revolutionary by design.

## Notes

1. For additional information on the JSTOR digital library please see; Roger C Schonfeld, *JSTOR: A History*. (Princeton: Princeton University Press, 2003), eBook Collection (EBSCOhost), EBSCOhost (last accessed February 18, 2015); see also <http://about.jstor.org/about>.
2. America: History and Life [CD-ROM], ABC-Clio.
3. Kevin M. Guthrie “Lessons from JSTOR: User Behavior and Faculty Attitudes,” *Journal of Library Administration* 36, no. 3 (03, 2002): 118, doi:10.1300/J111v36n03\_10.
4. William H. Walters, “E-books in Academic Libraries: Challenges for Discovery and Access,” *Serials Review* 39,2 (May, 2013): 97-104, for a discussion of problematic timing in publication of ebooks see Julia Proctor, “How Old is That Ebook: A Call for Standardization in Publisher-Provided Ebook Publication Dates,” *Journal Of Electronic Resources Librarianship* 25, no. 2 (April 2013): 100-114. Academic Search Premier, EBSCOhost (last accessed February 18, 2015). doi:10.1080/00987913.2013.10765501.
5. Donald A. Barclay “The Myth of Browsing,” *American Libraries* 41, no. 6/7 (June/July, 2010): 53.
6. Barclay, 54.
7. Linda Burfield Hazzard, *Scientific Fasting: the Ancient And Modern Key to Health*, 5th, rev. and amplified ed. of *Fasting for the cure of disease*. (New York: Grant Publications, 1927).
8. “Public Library is Abandoning Disputed Plan for Landmark,” *The New York Times* (New York, NY), May 7, 2014, last accessed February 18, 2015, [http://www.nytimes.com/2014/05/08/arts/design/public-library-abandons-plan-to-revamp-42nd-street-building.html?\\_r=0](http://www.nytimes.com/2014/05/08/arts/design/public-library-abandons-plan-to-revamp-42nd-street-building.html?_r=0) ; see also “In Renderings for a Library Landmark, Stacks of Questions,” *The New York Times* (New York, NY), January 29<sup>th</sup>, 2013, last accessed February 18, 2015, <http://www.nytimes.com/2013/01/30/arts/design/norman-fosters-public-library-will-need-structural-magic.html?pagewanted=all>.
9. Heather Christenson, “HathiTrust: A Research Library at Web Scale,” *Library Resources & Technical Services* 55, no. 2 (April 2011): 94-95. Academic Search Premier, EBSCOhost (last accessed February 18, 2015). see also Heather Christenson and John P. Wilkin, “Intellectual Property Rights & the HathiTrust Collection,” 3 (Paper presented at UNESCO’s 2012 conference The Memory of the World in the Digital Age,” Vancouver, BC, September, 2012), (last accessed February 18, 2015), [http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/VC\\_Christenson\\_Wilkin\\_26\\_A1330.pdf](http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/VC_Christenson_Wilkin_26_A1330.pdf).
10. Jeremy York, “A Preservation Infrastructure Built to Last; Preservation, Community, and HathiTrust,” 8 (Paper presented at UNESCO’s 2012 conference The Memory of the World in the Digital Age, Vancouver, BC, September, 2012), last accessed February 18, 2015, [http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/VC\\_York\\_26\\_B\\_1410.pdf](http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/VC_York_26_B_1410.pdf).
11. Chant, Ian. “Academic: Appeals Court Upholds HathiTrust Verdict,” *Library Journal* 139, no. 12 (July, 2014): 15. Academic Search Premier, EBSCOhost (last accessed February 18, 2015). See also “Unlocking the Riches of HathiTrust,” *American Libraries* 44, no. 1 (January, 2013): 40-43, last ac-

cessed 02-18-2015, <http://www.americanlibrariesmagazine.org/article/unlocking-riches-hathitrust>.

12. York, 1.
13. HathiTrust currently requires logging in as a partner institution in order to access the full-text, standalone, PDF file download on items digitized by Google. For additional information regarding current download limitations related to full text PDF files of items scanned by Google please see Help—Using the Digital Library, (last accessed February 18, 2015), [http://www.hathitrust.org/help\\_digital\\_library#Download](http://www.hathitrust.org/help_digital_library#Download).
14. For additional information on HathiTrust please see Sarah M. Pritchard, “HathiTrust Libraries Map a Shared Path: A Turning Point in Information Access,” *Portal* 12, no. 1 (2012): 1-3; Jeremy York and Brian E. C. Schottlaender, “The Universal Library is Us: Library Work at Scale in HathiTrust,” *Educause Review* 49, no. 3 (May, 2014): 48-49, Academic Search Premier, EBSCOhost (last accessed February 18, 2015), also see Papers and Presentations, (last accessed February 18, 2015), [http://www.hathitrust.org/papers\\_and\\_presentations](http://www.hathitrust.org/papers_and_presentations).