HathiTrust: Transforming the Library Landscape

By Diane Parr Walker

Many in attendance at the HathiTrust Constitutional Convention (Constitutional Convention 2011, n.d.) in October 2011 seemed to feel that the event and its outcomes were historic. Whether or not that was indeed the case will be determined in the future. For now, I will offer one perspective on the organization and on the collective actions taken during three days this past fall by representatives from the voting member institutions that made up the HathiTrust community at this moment in time.

What is HathiTrust?

In the words of John Wilkin, HathiTrust Executive Director and Associate University Librarian for Library Information Technology at the University of Michigan, "HathiTrust is a remarkable example of collective action, of our community working together to solve a common problem" (Wilkin, 2011).

The HathiTrust website describes the effort and its history:

> HathiTrust began in 2008 as a collaboration of the 13 universities of the Committee on Institutional Cooperation, the University of California system, and the University of Virginia to establish a repository to archive and share their digitized collections. HathiTrust has quickly expanded to include additional partners and to provide those partners with an easy means to archive their digital content.

> The initial focus of the partnership has been on preserving and providing access to digitized book and journal content from the partner library collections. This includes both in copyright and public domain materials digitized by Google, the Internet Archive, and Microsoft, as well as through in-house initiatives. The partners aim to build a comprehensive archive of published literature from around the world and develop shared strategies for managing and developing their digital and print holdings in a collaborative way.

> The primary community that HathiTrust serves are the members (faculty, students, and users) of its partners libraries, but the materials in HathiTrust are available to all to the extent permitted by law and contracts, providing the published record as a public good to users around the world (Our Partnership, n.d.).

Hathi (pronounced HAH-tee), is the Hindi word for elephant. The name is intended to underscore the immensity of this undertaking, but also to evoke memory, wisdom, and strength ("Major library partners," 2008). The original mission established for HathiTrust is "to contribute to the common good by collecting, organizing, preserving, communicating, and sharing the record of human knowledge" (Mission and Goals, n.d.). Its founding goals are:

1. To build a reliable and increasingly comprehensive digital archive of library materials converted from print that is co-owned and managed by a number of academic institutions.

2. To dramatically improve access to these materials in ways that, first and foremost, meet the needs of the co-owning institutions.

3. To help preserve these important human records by creating reliable and accessible electronic representations.

4. To stimulate redoubled efforts to

coordinate shared storage strategies among libraries, thus reducing long-term capital and operating costs of libraries associated with the storage and care of print collections.

5. To create and sustain this "public good" in a way that mitigates the problem of free-riders.

6. To create a technical framework that is simultaneously responsive to members through the centralized creation of functionality and sufficiently open to the creation of tools and services not created by the central organization.

Three libraries in Indiana are members of HathiTrust as of October 2011. Indiana University along with the University of Michigan led the formation of the initiative. Purdue University, as a member of the Committee on Institutional Cooperation (CIC), was also among the founding members. The University of Notre Dame joined in May 2011. Although only member institutions that joined before October 2010 were allowed to speak and vote at the Constitutional Convention, Notre Dame representatives, along with thirteen other new members, were invited to attend as observers. [1]

At the time of the Constitutional Convention, the HathiTrust website statistics reported cumulative digital holdings of nearly 10 million volumes, representing over 5 million book titles and 250,000 serial titles. In his remarks to the Convention, Wilkin stated that already by June 2010 (two years after the formation of HathiTrust) "nearly every ARL (Association of Research Libraries) library could depend on finding approximately 31% of its collection online in HathiTrust" (Wilkin, 2011). At current rates of ingestion to the database, Wilkin estimated that by early 2012 research libraries will likely be able to find about 50% of their collections in digital form in the archive, and he noted that rates of overlap between the HathiTrust digitized collections and those of individual academic libraries are even higher for non-ARL institutions.

Re-stated in perhaps overly-simplistic terms, HathiTrust was formed to preserve securely the digitized content of the printed holdings of research libraries, created through such programs as the Google Book Project (http://www.google.com/googlebooks/library.html), the Internet Archive (http://www.archive.org/), and individual library digitization projects, and to provide full text access for research and academic purposes.

If the oft-uttered assumption that in the future anything not digital won't be found and used is true, HathiTrust is the means of assuring that the record of past human intellectual effort, at least as represented by the collective holdings of U.S. research libraries, will not pass into non-digital obscurity. That alone is a significant accomplishment, but is not all the HathiTrust partnership has accomplished so far or aspires to accomplish in the future.

Significant Accomplishments before the Constitutional Convention

In his opening presentation at the convention, Wilkin emphasized that the HathiTrust partnership has gone beyond the initial goal of a collective repository to create a "rich, open system with a nuanced understanding of rights and the ability to deliver various forms of content to different audiences in different ways" (Wilkin, 2011). His emphasis, though, was on the fact that this has been accomplished collectively, through the contributed efforts of the partner/member[2] institutions rather than by an external organization for libraries.

Wilkin highlighted three significant accomplishments, among the many the partnership can site (see others at http://www.hathitrust.org/news_publications).

- The HathiTrust has been certified by the Center for Research Libraries (CRL) as a trustworthy digital repository. CRL established a rigorous Trustworthy Repositories Audit and Certification (TRAC) assessment program in 2007 (Center for Research Libraries, n.d.). TRAC Certification requires a rigorous

and thorough review by an advisory panel of leaders in library collection development, administration, and digital technology, based on a set of criteria intended to "identify digital repositories capable of reliably storing, migrating, and providing access to digital collections" (Center for Research Libraries, 2007, p. Foreword). In other words, this certification signifies that the permanence of HathiTrust digital content can be relied upon in ways that ongoing accessibility of much web content cannot be.

- HathiTrust has created a viable full-text search mechanism that works across all of the content in the repository. In Wilkins's words: "I hope no one here is so jaded as to think that full-text searching across millions of volumes is a slam-dunk… Our users can now search over 3 billion words and get results in a split second. Collective work in the partnership has produced faceted results in our full text, and ranking that takes bibliographic information in the full text into account. The functionality that we have today is tremendous, and it provides a foundation for a next generation of search that gives our users access to bibliographic information where needed, and full text where desired (Wilkin, 2011).

Indiana University and the University of Illinois have created a research center to support uses of the HathiTrust collection (http://newsinfo.iu.edu/news/page/normal/19252.html). The intent of this initiative is to develop methods of using computational tools to analyze texts across the HathiTrust collection in ways that do not infringe copyright or require that the researcher read individual items in the collection. As Wilkin observed, the significance of this development is that the research center initiative "comes from faculty leadership (from those who would *do* the research), drawn to use of this immense library, rather than from us [i.e., libraries] in support of those faculty" (Wilkin, 2011).

In addition, HathiTrust has begun working with publishers and authors to make it possible for rights holders to open access to their works in HathiTrust and for publishers to deposit digital master files for perpetual archiving. Also, the collective is beginning to facilitate lawful uses of in-copyright materials, including access for users with print disabilities and developing collaborative methods for identifying and providing access to orphan works.

Constitutional Convention

Wilkin observed in his opening remarks at the Constitutional Convention that since beginning large-scale digitization of library collections in 2005 through such efforts as the Google Book Project and the Internet Archive, the research library community has digitized over half of the collective holdings of ARL libraries. The launching of HathiTrust in 2008 established an organization to facilitate collective action among research libraries on a grand scale. He asserted that "seldom has so much in our world changed in such a short time. Together, we have utterly transformed parts of the library landscape" (Wilkin, 2011).

Why hold a Constitutional Convention? As articulated on the HathiTrust website:

HathiTrust was chartered by the founding partners for an initial five-year period, from 2008-2012, with a formal review of governance and sustainability to be conducted by the partners in the third year. In October, 2011, institutions who joined HathiTrust by October 31, 2010 will participate in a Constitutional Convention to determine the governance model for the partnership and set directions for its next phase (Governance, n.d.).

HathiTrust contracted with Ithaka S+R to conduct the required three-year review of HathiTrust's accomplishments and progress. The resulting report provided background for the Convention (Ithaka S+R, 2011).

The review focused on three inquiries:

1. How participating libraries perceive the value of HathiTrust

2. Expectations of participating libraries for operating and governing the initiative moving forward

3. The views of libraries that do not yet participate in HathiTrust

Findings revealed that among participating libraries HathiTrust is valued as a preservation service, that participating libraries value the fact that HathiTrust is a collaborative effort "growing from, and led by the research community," (Demas, 2011, p.9) and that there is significant goodwill and enthusiasm for the HathiTrust initiative and its missions. Among the questions raised was whether HathiTrust has the potential to help manage print collections (Demas, 2011, p. 9).

This set of findings seemed to give significant imperative to continuing to move forward as a partnership, rather than moving toward alternative models such as forming a separate non-profit enterprise or contracting with an external party for services. Ithaka S&R concluded that "the current structure of HathiTrust is somewhat unique in that it is a collaboration that is trying to deeply operationalize tasks in a distributed, volunteer mode. This is an opportunity to investigate collaborative staffing models and is perhaps a key role for HathiTrust with its innovative spirit" (Demas, 2011, p. 12).

Regarding expectations, those interviewed for the review noted the high amount of expertise and effort contributed by partner institutions but questioned whether it is sustainable to build an organization based on "volunteer" effort. Consultation, input, and transparency emerged as core desires for governance going forward, along with a general sense that those partners contributing more in terms of effort and content should rightfully have more "weight" in decision-making.

As for views of libraries not currently members of HathiTrust, there was strong interest in the mission and preservation function of HathiTrust from libraries with rare and unique collections, but a perception that there is currently not an effective ingest method for these sorts of

materials. International libraries are hesitant to join an initiative perceived as U.S.-based. Some regional and consortial digital library projects see HathiTrust as duplicative (Demas, 2011, p. 10).

While the report of the review provided substantial documentation on how HathiTrust is perceived by member and non-member institutions and a summary of areas to which it was suggested HathiTrust might next turn its attention, the Convention was structured around ballot proposals submitted by partner institutions in advance. The structure was relatively formal, with sessions allotted to consider each of seven ballot proposals through presentation of the proposal, invitation and discussion of amendments to the proposal, and a vote by the partnering institutions and consortia that had joined HathiTrust prior to October 31, 2010. Voting was weighted to allocate a certain number of votes to each participating institution, according to a formula that considered each member's relative financial and digital content contribution. The list of ballot proposals and the outcome of voting was posted to the HathiTrust website almost immediately (Constitutional Convention Ballot Proposals, n.d.).

By passage of the ballot proposal to establish an effective governance structure, the convention determined that at the end of HathiTrust's initial five-years (2008-2012), the partnership will be led by a Board of Governors, a Board Executive Committee, and Board-appointed committees "to ensure timely review of current issues, including periodic review of the cost model" and that "HathiTrust's governance [shall] be based on a set of clearly articulated Bylaws and a process for amending them" (Proposal 3 - Governance Structure, n.d.). It is important to understand that Bylaws do not yet exist, nor does the framework for electing the Board of Governors. The ballot proposal resolves that the Board of Governors shall develop Bylaws to be presented to the membership for ratification or rejection within six months of its formation. Passage of this ballot proposal presumably sets in motion action on the part of the

founding Executive Committee to implement formation of and election to the Board of Governors in the coming months. Overall, passage of this proposal represented a strong vote of confidence in the current executive management.

The other ballot proposals passed chart a new and ambitious expansion of HathiTrust's goals. Based on those proposals, we can look to this partnership to undertake several specific activities:

- Establish a distributed print monograph archiving program among HathiTrust member libraries (Proposal 1 - Distributed Print Monographs Archive, n.d.).

- Formalize a transparent process for inviting, evaluating, ranking, launching and assessing development initiatives (including the incorporation of existing services and tools along with those to be newly developed) from HathiTrust partner institutions (Proposal 2 - Approval Process for Development Initiatives, n.d.).

- Through coordinated and collective action, expand and enhance access to U.S. Federal publications, including those issued by the Government Printing Office and other federal agencies (Proposal 4 - U.S. Government Documents, nd.).

- Develop and vet a fee-for-service model to allow contribution of content from non-partner entities (Proposal 7 - Fee-for-service Content Deposit, n.d.).

A proposal to create a HathiTrust committee to review the implications and applicability of service, policies, and procedures developed by its members before they are adopted generated much discussion (Proposal 6 - HathiTrust Implementation Review Committee, n.d.) but was not passed. According to the rules established for the Constitutional Convention, the proposal will likely not be taken up again.

One remaining proposal (Proposal 5 - Mission and Goals, n.d.), to broaden the stated mission of HathiTrust to expand its focus from building "a digital archive of library materials converted from print" to considering the full range of "digital assets of intellectual value" to researchers, also generated much discussion. In the end, the Convention voted to refer this proposal to the incoming Board of Governors for further consideration.

Does the HathiTrust and this Constitutional Convention transform the library landscape? I would say that both are a significant part of what Abby Smith Rumsey, who convened and moderated the Convention, described as "a deliberate transition to 21st century modes of scholarly communication" (Demas, 2011, p. 5). In an open letter regarding Ithaka S+R's three-year review, Ed Van Gemert, Chair of the HathiTrust Strategic Advisory Board, observed, "The strength of the collaborative is in thinking big while engaging in focused strategic work. The challenge is to remain focused on strategic needs by not trying to be everything for everyone, while at the same time continuing to tackle large issues" (Van Gemert, n.d.). The HathiTrust Constitutional Convention aimed to do just that: to think big while focusing on the strategic need to establish a long-term governance structure to ensure future progress. The ballot proposals that were passed, together with the long list of future potential activities and issues to be addressed that emerged during discussion (Demas, 2011, pp. 3, 36-38), will test the ongoing strength of the collaborative. Accomplishments to date, in only three years, and the evident commitment of the participants in the Convention bode well for the future. This collective commitment "to contribute to the common good by collecting, organizing, preserving, communicating, and sharing the record of human knowledge" in digital and digitized form is changing the landscape of libraries (Mission and Goals. (n.d.). Through HathiTrust, libraries are pooling resources and mobilizing expertise at an unprecedented scale to address common problems, ranging from preserving and accessing the collective content in research libraries, to beginning to approach questions

of copyright and fair use in a progressive and coordinated way. The Constitutional Convention confirmed the collective commitment of the member institutions to aggressively and intentionally change our landscape.

Reference

Center for Research Libraries. (n.d.). *Certification and assessment of digital repositories*. Retrieved from http://www.crl.edu/archiving-preservation/digital-archives/certification-and-assessment-digital-repositories

Center for Research Libraries. (2007). *Trustworthy repositories audit & certification: Criteria and checklist.* Chicago: Center for Research Libraries and OCLC Online Computer Library Center, Inc.

Constitutional Convention 2011. (n.d.). Retrieved from http://www.hathitrust.org/constitutional_convention2011

Constitutional Convention Ballot Proposals. (n.d.). Retrieved from http://www.hathitrust.org/constitutional_convention2011_ballot_proposals

Demas, S. (2011, October 8-9). *Notes of the HathiTrust Constitutional Convention*. Retrieved from http://www.hathitrust.org/documents/HathiTrust-ConCon-Notes.pdf

Governance. (n.d.). Retrieved from http://www.hathitrust.org/governance

Ithaka S+R. (2011). Briefing paper on progress and opportunities for HathiTrust. Retrieved from http://www.hathitrust.org/documents/hathitrust-3year-review-2011.pdf

Major library partners launch HathiTrust shared digital repository. (2008, October 13). Retrieved from http://ns.umich.edu/new/releases/6774

Mission and Goals. (n.d.). Retrieved from http://www.hathitrust.org/mission_goals

Our Partnership. (n.d.). Retrieved from http://www.hathitrust.org/partnership

Proposal 1 - Distributed Print Monographs Archive. (n.d.). Retrieved from http://www.hathitrust.org/constitutional_convention2011_ballot_proposals#proposal1

Proposal 2 - Approval Process for Development Initiatives. (n.d.). Retrieved from http://www.hathitrust.org/constitutional_convention2011_ballot_proposals#proposal2

Proposal 3 - Governance Structure. (n.d.). Retrieved from http://www.hathitrust.org/constitutional_convention2011_ballot_proposals#proposal3

Proposal 4 - U.S. Government Documents. (n.d.). Retrieved from http://www.hathitrust.org/constitutional_convention2011_ballot_proposals#proposal4

Proposal 5 - Mission and Goals. (n.d.). Retrieved from http://www.hathitrust.org/constitutional_convention2011_ballot_proposals#proposal5

Proposal 6 - HathiTrust Implementation Review Committee. (n.d.). Retrieved from http://www.hathitrust.org/constitutional_convention2011_ballot_proposals#proposal6

Proposal 7 - Fee-for-service Content Deposit. (n.d.). Retrieved from http://www.hathitrust.org/constitutional_convention2011_ballot_proposals#proposal7

Van Gemert, E. (n.d.). *Open letter regarding Ithaka S+R review of HathiTrust*. Retrieved from http://www.hathitrust.org/constitutional_convention2011

Wilkin, J. (2011, October). *HathiTrust's Past, Present, and Future*. Retrieved from http://www.hathitrust.org/blogs/perspectives-from-hathitrust/hathitrusts-

past-present-and-future

Footnotes

1. Prior to joining the University of Notre Dame in July 2011, I was Deputy University Librarian at the University of Virginia, an early partner in the HathiTrust.

2. The terms "partner" and "member" are used interchangeably in HathiTrust documentation and during discussions at the Constitutional Convention.

BIO

Diane Parr Walker, Edward H. Arnold University Librarian
Hesburgh Libraries University of Notre Dame
Email: dwalker6@nd.edu

Diane Parr Walker joined the University of Notre Dame as Edward H. Arnold University Librarian in July 2011. Prior to that, she was the Deputy University Librarian at the University of Virginia.