

**Briefing Paper on Progress and Opportunities for HathiTrust  
Ithaka S+R for the HathiTrust Strategic Advisory Board  
15 July 2011**

**Executive Summary**

The Strategic Advisory Board for HathiTrust, a library-based initiative to provide preservation and access for library content digitized by Google and other parties, is preparing for a Constitutional Convention in October 2011. At the Convention, representatives from the libraries that participate in HathiTrust will gather to discuss the future direction of the initiative.

To help frame the agenda for the Convention, Ithaka S+R has conducted research on the attitudes and needs of existing HathiTrust partner libraries and libraries that do not yet participate, to help the Board frame the most important issues for the October meeting. Drawing on a survey of HathiTrust participating libraries, interviews with HathiTrust participating libraries and libraries that do not yet participate in HathiTrust, and other research, we see the following as important next steps and Convention discussion items for HathiTrust:

- *Clearly defining objectives for the next 3-5 years, possibly mapping out the rationale for those objectives in the context of a revised mission statement, and enhancing the pipeline of information about HathiTrust's strategic priorities to partner libraries.* Respondents to a survey of HathiTrust partner libraries and follow-up interviewees spoke in particular about the need for clear communication from the HathiTrust staff and leadership, noting in several cases that they have not always felt well apprised of contemplated new initiatives. To the extent that HathiTrust is able to provide a more constant flow of information about proposed initiatives, that would seem to be valued by partner libraries.
- *Advancing changes that allow for nomination and selection to the Executive Committee, and clearly communicating the options to the HathiTrust partner community ahead of the Constitutional Convention.* A number of libraries, particularly those that have contributed content to the archive, noted that they expect heavy content contributors and/or founding members to have a more heavily weighted voice in governance under any future model.
- *Demonstrating to partner libraries the sustainability and feasibility of the new cost model for HathiTrust.* Some current participants raised concern about the increasing costs of HathiTrust, and libraries that do not yet participate worried about a perceived instability in the pricing structure. If HathiTrust elects to move forward with a version of the proposed cost model, there may be value in introducing a stabilizer into the cost model—either asking that institutions commit to membership for a set period, or calculating the divisor in the formula for public domain volumes (the number of institutions) using a rolling average that takes into account anticipated uptake in the next year.

## Introduction

In October 2011, representatives from the partner libraries that make up HathiTrust (<http://www.hathitrust.org>) will gather for a Constitutional Convention, where they will reflect on the progress of the past three years, discuss potential future directions for HathiTrust, and vote on a series of ballot initiatives related to governance, administration, and finance. The Strategic Advisory Board for HathiTrust, a group of library leaders who oversee the strategic direction for the initiative, engaged Ithaka S+R to prepare a briefing paper to help the Board and HathiTrust's other committees and working groups to frame the broad discussion for the Convention.

As part of our working approach for the project, we sought to weigh respondents' feedback and the opportunities that they presented against HathiTrust's broader mission "to contribute to the common good by collecting, organizing, preserving, communicating, and sharing the record of human knowledge," and also against its six subsidiary goals:

- 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.
- To dramatically improve access to these materials in ways that, first and foremost, meet the needs of the co-owning institutions.
- To help preserve these important human records by creating reliable and accessible electronic representations.
- 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.
- To create and sustain this "public good" in a way that mitigates the problem of free-riders.
- 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.<sup>1</sup>

How do libraries' stated perceptions of, and needs from, HathiTrust converge with these goals, and where do they diverge?

In the sections that follow, we draw on our research to address three angles of input from the library community:

- (1) *How participating libraries perceive the value of HathiTrust.*
- (2) *Participating libraries' expectations for operating and governing the initiative moving forward.*
- (3) *The views of libraries that do not yet participate in HathiTrust.*

---

<sup>1</sup> [http://www.hathitrust.org/mission\\_goals](http://www.hathitrust.org/mission_goals)

## Methodology

Ithaka S+R carried out several types of research. First, we administered a web-based survey to the 52 libraries that participated in HathiTrust as of 15 May 2011. The survey was open from 18 May to 3 June, and received 46 responses, for a response rate of 88%. The survey asked respondents a series of attitudinal and factual questions related to HathiTrust, to other preservation and access platforms that the partner libraries may use, and to the libraries' internal policies and spending on digital preservation. The survey was emailed to each library's designated representative for HathiTrust, with a request to forward the survey to the library's director or to someone who could speak to the library's overall perspective on participating in HathiTrust.

We supplemented the survey results with a series of follow-up calls with eight survey respondents, both to clarify their initial responses and to ask follow-up questions. To broaden our understanding of the academic and research library community's perceptions of, and needs from, HathiTrust, we sought interviews with representatives from 25 libraries that do not yet participate in HathiTrust, of which 14 consented to an interview.<sup>2</sup> To get an anecdotal perspective of the ways in which the access and research functions of HathiTrust might be made more useful to scholars, we sought interviews with 15 scholars and subject librarians in history, language and literature, linguistics, and information science, of whom five consented to speak with us. (Although these interviews informed our research approach, they are not discussed directly in this paper.) Finally, throughout the project, our meetings and interviews with HathiTrust stakeholders, including members of the Strategic Advisory Board and the Executive Director, provided a helpful internal perspective.

## Value to Participating Libraries

In the survey and follow-up interviews, participating libraries strongly voiced their support for HathiTrust, citing its value as a digital preservation service, as an alternative to what more than one respondent called "the Google monopoly," and as an initiative growing from, and led by, the research library community. At a high level, there is clearly a great deal of goodwill for the initiative and its mission.

### *How do libraries rate HathiTrust's progress in meeting short-term functional objectives?*

As part of the research process, the Strategic Advisory Board requested that Ithaka S+R use an evaluation of HathiTrust's functional objectives (<http://www.hathitrust.org/objectives>), a list of developments undertaken by the HathiTrust community, as one guidepost for our research. Respondents to the survey were asked to indicate their agreement with a series of statements related to the functional objectives, using a scale from 1 to 10 (where 1 represented strong disagreement with the statement in question and 10, strong agreement).

The average and median responses showed overall satisfaction with the quality of the image scans in HathiTrust and the Trust's capacity for preserving and migrating content. Respondents

---

<sup>2</sup> In general, libraries outside the United States were less responsive to our requests for an interview.

gave slightly lower marks to the HathiTrust user interface and the branding/watermarking of items contributed by member libraries, although the median (5) for each of these two statements falls in the middle of the 1-10 satisfaction range.

*Table One.* Satisfaction with HathiTrust’s progress on functional objectives

Survey Statement	Average	Median
HathiTrust is well positioned to migrate the content that my library has contributed to new formats as the need arises.	6.98	7
The quality of scans presented in HathiTrust is generally high.	6.91	7
Compared with my library working on its own, HathiTrust would be more efficiently and effectively able to migrate the content that my library has contributed to new formats.	7.12	8
We are satisfied with the prominence and placement of our organization’s branding on individual items that we have contributed to HathiTrust.	5.63	5
The HathiTrust public user interface meets all the needs of our users.	4.74	5

*Branding.* Participating libraries raised two concerns about the local branding of their materials in HathiTrust. First, several respondents voiced a desire for a more prominent marking on the materials that they have contributed to HathiTrust than the marks of ownership that appear on materials now—one respondent called the existing watermarking arrangement “pretty subtle,” and another reported that they would like the materials they contribute to carry their institution’s logo or something more visually attractive than plain text.

Second, a small number of participating libraries indicated that they expect their branding to appear on the scans of all items that they have contributed, even if the version being presented to the user was not itself digitized using the library’s copy. As one respondent put it: “If [an item in HathiTrust] remains primarily branded by whichever institutions contributed the display copy, it [HathiTrust] will not give us the promotional edge we need.” As libraries further transition from local to networked models for collections and services, they may become more comfortable giving up control over the placement of their name in services like HathiTrust; however, “at this point in time,” one respondent wrote, “it is important to make the institutional branding [in HathiTrust] more prominent in order to help us to justify participation.”

*User interface.* Respondents pointed to a number of incremental fixes to the HathiTrust user interface, noting in particular the importance of reflecting serial holdings accurately in search results: two respondents noted that individual serial volumes need to be indexed correctly,

particularly for those serials that were bound as a single volume by a contributing library and are now reflected in HathiTrust as one item.

*To what extent does participation in HathiTrust help libraries to realize cost savings or cost avoidance?*

*Digital storage.* One potential value for joining HathiTrust is the ability to free up resources (funds, staff time, local server space) that would be devoted to local storage of digitized content and divert them to other functions. The survey asked participating libraries to indicate whether participating in HathiTrust caused them to change their spending on storing, managing, and preserving digital assets using their institution's local infrastructure, and if so, by what dollar amount. Of the 25 participants that reported having contributed content to HathiTrust, only five reported having changed their local spending as a result of HathiTrust. Of those five:

- Two reported increasing local spending; in both cases the respondents cited the staff time associated with preparing HathiTrust files for ingest as the driver.
- Three reported decreasing their local spending, although only one was able to provide an estimate of those savings (\$150,000).

Although the survey did not prompt libraries to estimate cost avoidance, several libraries noted in free response boxes within the survey that they prepared projections comparing the costs of storing digital content locally with the cost of joining HathiTrust before deciding to join the initiative.

A number of respondents made clear in follow-up interviews that participation in HathiTrust is not simply a question of money—the overarching mission, collaboration among the academic and research library communities, and the chance to help prioritize developments for the archive are all important drivers for participation. With that said, the question of how HathiTrust impacts local storage costs seems important because the initiative is considering moving to a new cost model; the model, which is outlined on the HathiTrust website, uses a formula that apportions a fraction of the storage cost for a digitized volume to all partner libraries who have owned a print copy of that volume.<sup>3</sup> As HathiTrust grows and sets new directions, the proposed cost model allows for libraries' participation fees to increase in a way that was unlikely under the existing cost model (which was based solely on the cost per GB of digitized content, a cost that libraries could control): for example, if a library were to join HathiTrust with a significant amount of public domain content, all partner libraries could see an increase in fees. In a period of challenged library budgets, the ability of a service like HathiTrust to demonstrate cost savings for libraries—particularly with the potential for an increase in membership fees—will likely become more important.

*Print collections management.* Another potential source of cost savings or avoidance for HathiTrust participants is the use of the digital archive as a tool for managing print collections, as libraries face the familiar problem of space constraints. However, Ithaca S+R's past research

---

<sup>3</sup> <http://www.hathitrust.org/cost>

has shown relatively cautious interest among in the U.S. higher education community about the speed of a print-to-electronic transition for books: the Ithaka S+R Faculty Survey 2009 found that only 4% of respondents agreed with the statement “Within the next five years, the use of e-books will be so prominent among faculty and students that it will not be necessary to maintain local collections of hard-copy books”; in an Ithaka S+R survey of U.S. academic library directors conducted in 2010, just 7% of respondents agreed with a similar statement.<sup>4</sup>

Those relatively low responses may shed light on why HathiTrust participating libraries seem not to be using HathiTrust as an integral part of their print collections management strategy. While slightly more than half of the responding institutions reported that the existence of a digitized copy in HathiTrust is one factor that they consider in deciding whether to move or dispose of books, no library reported that the existence of a digitized copy in HathiTrust is a prerequisite for those decisions. (Anecdotally, follow-up interviewees reported Google Books, Internet Archive, and other libraries’ repositories as other acceptable sources for a digitized copy when making print collections management decisions.) Similarly, only two libraries reported that the existence of digitized copies in HathiTrust has had an impact on the rate at which they move print books to off-site storage, and only six reported that HathiTrust has an impact on the rate at which they dispose of print books.

*Table Two.* HathiTrust’s importance in print collections management decisions

Impact of HathiTrust on Decision-Making	Decision to move print books off-site		Decision to dispose of print books	
	Content contributors	Other members	Content contributors	Other members
Prerequisite <sup>5</sup>	-	-	-	-
One Factor	13	12	12	12
Not a Factor	7	7	8	8
N/A	5	2	5	1

<sup>4</sup> See: Roger C. Schonfeld and Ross Housewright, *Ithaka S+R Faculty Survey 2009: Key Strategic Insights for Libraries, Publishers, and Societies* (7 April 2010), p. 23, <http://www.ithaka.org/ithaka-s-r/research/faculty-surveys-2000-2009/Faculty%20Study%202009.pdf>; Matthew P. Long and Roger C. Schonfeld, *Ithaka S+R Library Survey 2010: Insights from U.S. Academic Library Directors* (5 April 2011), p. 36, <http://www.ithaka.org/ithaka-s-r/research/ithaka-s-r-library-survey-2010/insights-from-us-academic-library-directors.pdf>.

<sup>5</sup> The “Prerequisite” row in this chart captures two response choices in the survey: “The existence of a digitized copy in HathiTrust that is immediately available for viewing by our users is a prerequisite” and “The existence of a digitized copy in HathiTrust, whether or not it is immediately available for viewing by our users is a prerequisite.”

*Table Three. Impact of HathiTrust on rate of changes to print collections<sup>6</sup>*

<u>Impact of HathiTrust on Rate</u>	<u>Moving print books off-site</u>		<u>Disposing of print books</u>	
	<u>Content contributors</u>	<u>Other members</u>	<u>Content contributors</u>	<u>Other members</u>
Increased	1	1	1	-
Maintained	-	-	2	3
Decreased	-	-	-	-
No Impact <sup>7</sup>	19	17	16	15
N/A	5	2	5	1

This is not to say that HathiTrust does not have a continued and vital role to play in reducing the costs associated with managing print collections. In the aforementioned Ithaka S+R Library Survey 2010, U.S. academic library directors were presented with the following concept statement (modeled closely on HathiTrust):

Suppose there existed a robust system for preservation of and access to historical monograph collections. In such a system, the millions of books digitized through the Google library digitization project would be readily available in digital form for use by your community. Digital preservation would be assured by a trusted third party archive, and the paper source materials would be preserved in a suitable number of print archives. Ultimately, discovery and accessibility would be greatly improved while preservation would be assured.

When asked how the existence of such a system might impact their print collections management strategies, 74% of respondents said that withdrawal of print books would be an important strategy for their libraries, and 84% of respondents reported that they would be more likely to withdraw their print book collections under such a scenario if their library faced significant space pressures.<sup>8</sup> In addition, 90% of respondents indicated that they would be likely to withdraw their print book collections if their library could provide guaranteed on-demand access to the print versions through a sharing network. Clearly, there is interest in the library community in using a system like HathiTrust for print collections management, and it may be that either an improvement in the discoverability and accessibility of titles in HathiTrust, the addition of a print preservation and/or access service, or the simple passage of time, will nudge libraries toward greater use of HathiTrust for print collections management.

<sup>6</sup> It bears particular noting that these figures were self-reported by libraries.

<sup>7</sup> The distinction between "Maintained" and "No Impact" was meant here to signify that a library could conceivably choose to maintain, rather than decrease, the rate at which they move or dispose of print books because of the availability of digitized copies of those works in HathiTrust.

<sup>8</sup> Long and Schonfeld (2011), pp. 36-37.

### **Future Operations for the Initiative**

HathiTrust's existing operational structure is designed to leverage cooperative work by its network of partner libraries: the work of a relatively modest central staff at the University of Michigan is bolstered by volunteered time from partner library representatives on working groups and governing committees. In addition, a smaller set of partner libraries are devoting more concentrated time to the development of tools for the benefit of the larger membership, as in the case of the California Digital Library's extensive work on metadata management and the collaborative development of the Research Center by Indiana University and the University of Illinois, Urbana-Champaign, which will allow researchers to conduct text mining across the HathiTrust corpus. The overall model of contributed staff time from partner libraries clearly appealed to the librarians interviewed and surveyed for the project—one interviewee called participation on a HathiTrust working group “one of the most satisfying experiences of my professional career.”

The section below is largely reflective of the views of the survey respondents and interviewees, and so it most directly speaks to HathiTrust's existing organizational model. With that said, there remains an overall question about whether a structure that depends heavily on contributed staffing and the buying-out of staff time from partner libraries will be sustainable over the long term. Contributed staff time may be retrenched by partner libraries during difficult budget periods; even in those cases where HathiTrust buys out the time of developers and project librarians at member institutions, the availability of that time may be curtailed as the partner institution's priorities change. (And as one interviewee pointed out, there is not, for example, an infinite supply of developers employed by research libraries for HathiTrust to draw on.) Evaluating whether and to what extent staff positions might need to be brought into the central HathiTrust organization would be difficult to forecast (and beyond the scope of this project), but HathiTrust's leaders and governing organizations will need to keep a pulse on staffing needs and be willing to spend the organization's surplus to support added internal staffing if the need arises.

*Partner libraries' expectations about staffing.* HathiTrust's existing constellation of governance committees and working groups are made up of library leaders whose time is contributed by HathiTrust partner institutions. In addition, and as mentioned above, a smaller number of partner libraries have taken on technological development work for the initiative. On one hand, the decentralized and devolved committee structure allows the HathiTrust community to benefit from the expertise of high-level library leaders with a wide range of competencies; on the other hand, the volunteer-centric staffing structure means that the continued vitality of the initiative depends heavily on the willingness of partner libraries to continue making these contributions.

In the survey administered to the HathiTrust community, partner libraries almost universally reported contributing staff time to the initiative: 43 of the 46 responding partner libraries responded that they devote some level of staff time to HathiTrust; of those, 32 reported a relatively modest allocation of staff time, between 0.1 and 1.0 FTE.



*Table Four. Projected Change in Contributed Staff Time<sup>9</sup>*

<u>Current Staffing Level</u>	<u>Increase</u>	<u>Maintain</u>	<u>Decrease</u>
None	1	2	-
0.1 to 1.0 FTE	17	15	-
1.1 to 2.0 FTE	2	3	-
2.1 to 3.0 FTE	2	1	-
3.1 to 4.0 FTE	1	-	-
More than 4.0 FTE	-	1	-

When partner libraries were asked to forecast how they expect their staffing contributions to HathiTrust to change (if at all) over the next 3-5 years, a roughly equal number reported that they expect to increase their staffing contributions, or maintain their current level (with no increase or decrease). No partner library contemplated decreasing or eliminating staffing contributions to HathiTrust.

*For what functions might libraries expect to contribute staff time in the future?*

Even as the partner libraries give a clear indication that they expect to maintain or increase the quantity of staff time to HathiTrust, planning for the initiative will need to take into account the qualities of those staffing contributions. One indicator for future staffing needs is the extent to which partner libraries prefer that certain functions for HathiTrust be carried out at their local institution (that is, using their own staff and resources, as in the case of contributed staff time to HathiTrust). The following table summarizes partner libraries' responses to a series of questions about their preferences for performing a broad set of functions related to HathiTrust locally:

*Table Five. Preferences for HathiTrust Functions: At the local institution or centralized?  
(On a scale from 1 to 10, where 1 = Local and 10 = Centralized)*

<u>Function</u>	<u>Average</u>	<u>Median</u>
Digital library infrastructure management	5.77	6
Digitization	4.51	4
Outreach and marketing	5.11	5
Research support	7.04	8
Rights determination	7.20	7.5
Tool development	7.11	8
User support	5.11	5

Perhaps unsurprisingly, libraries expressed a preference for functions that require technology development time, like tool development and the formation of research support services, to be accomplished by the central HathiTrust organization; tech development time is costly and likely scarce at those partner libraries that have developers on staff. Rights determination emerged as another function that partner libraries prefer to push up to the central HathiTrust

<sup>9</sup> The survey questions addressing current and projected contributions of staff time instructed respondents to exclude staff time devoted to general digitization activities.

organization (a response that is, perhaps, itself an endorsement of great value of the HathiTrust staff's research on orphan works and other rights issues).

The continued ability of the HathiTrust's leaders and governors to secure the staff time needed for these more complex functions will be crucial to the development of the initiative. If it becomes apparent that HathiTrust partner libraries are not able or willing to contribute staff time at the level needed to develop the project's infrastructure and interface, then the HathiTrust fee structure would need to be revised to reflect these new staffing needs.

*Expectations about representation in governance.* In the survey responses and follow-up interviews, several respondents voiced a desire for a method of consultation and input into the governing committees, including the Executive Committee. Respondents and interviewees noted that, under the initial HathiTrust governance model, there is not yet a process for nominating and selecting representatives to serve on the Executive Committee; they pointed to a process for doing so as a clear priority for the initiative.

With that said, interviewees had different views on the question of who should be eligible to serve on the governing committees: several respondents and interviewees from libraries that have contributed content posited that representation on the Executive Committee might be reserved for those libraries that have contributed content or have devoted significant staff time to the initiative; this has the virtue of rewarding those foundational libraries, but is also clearly at odds with the interests of libraries that have not contributed content but chose to join in part to have input into HathiTrust's strategic direction. (Ten of the 21 libraries in the survey who reported not having contributed content to HathiTrust also noted the opportunity to have input into the strategic direction of HathiTrust as one of their top three reasons for joining.)

If HathiTrust continues to operate under a membership governance model, the method for selecting representation may need to fall between two poles: on the one side, recognizing the contributions made by the heaviest contributors and/or founding members of the initiative, and on the other side, preserving some representation for those libraries with smaller collections that might wish to join HathiTrust in the future. One possible mechanism is a weighted voting system that gives those institutions that have contributed content and fulfill certain other parameters a greater share of voting power; another possibility would be a bicameral governance structure in which (for example) the upper committee would be reserved for founding members and heavy contributors of staff time, and the lower committee would be more broadly representative of the HathiTrust partner community.

*Desire for a clear communication of strategic objectives.* In the survey, respondents' questions about a model for governance and input moving forward are closely tied to confusion about the future priorities for HathiTrust: partner library representatives questioned both how HathiTrust will be shaped in the future, and how those changes will be communicated to the membership. "There is a lot of communication about what's happening already in HathiTrust, but not a lot about what's being contemplated," one participant noted. "For example, the development of the new cost model is something that we wish we could have been asked about: is this being

presented [at the Constitutional Convention] as something that we're thinking about, or as something that we definitely are using?" Although the articulation of a longer-term strategy for HathiTrust is one likely outcome of the Constitutional Convention and will itself serve as a useful opportunity to communicate with the HathiTrust community, the future governance structure for HathiTrust will need to include mechanisms for disseminating news of HathiTrust's priorities to the community, and for soliciting input in return.

*Intersection with other digital library initiatives.* Several respondents to the survey asked how HathiTrust will interact with the Digital Public Library of America (DPLA) initiative.<sup>10</sup> (Indeed, one scholar-interviewee who had not previously been aware of HathiTrust asked whether it was connected in some way with the DPLA.) To the extent that it will be possible at the Constitutional Convention, the Strategic Advisory Board and the Executive Committee should seek to clarify whether and how HathiTrust will interact with that initiative (itself still being defined).

### **Value to Potential Partner Libraries**

As detailed in the Methodology section, Ithaka S+R spoke with representatives from a range of libraries that did not yet participate in HathiTrust, including six ARL institutions, three U.S. liberal arts colleges, and five international libraries outside North America. These include a roughly even number of institutional interviewees who felt they would have content to contribute to HathiTrust, and those who do not currently contemplate engaging in digitization beyond small experiments. All of the interviewees expressed great interest in HathiTrust, and most reported that they had seriously contemplated joining HathiTrust over the past two years. (One of these libraries chose to join HathiTrust during the course of this research.) Their reasons for not joining provide useful context as the HathiTrust community considers shifting models for governance and finance, each of which will have implications for the ability of the initiative to attract new members.

In rough terms, the libraries that have not yet joined can be split into three categories:

*Libraries that only have interest in contributing special collections.* These institutions may have a small or nascent digitization program focused on rare and unique collections, and see a clear value in aggregating those materials with the larger HathiTrust corpus.

- These libraries have a strong interest in the mission and preservation function of HathiTrust, and possibly in using HathiTrust as a factor in making print collections management decisions.
- Among our interviewees, those at libraries that have examined HathiTrust in the past voiced a perception that the initiative does not yet have an effective ingest mechanism for special

---

<sup>10</sup> A useful summary article on the planned Digital Public Library of America is here: Gautam S. Kumar and Sirui Li, "Harvard Leads Digital Library Initiative," *Harvard Crimson* (12 April 2011). <http://www.thecrimson.com/article/2011/4/12/project-committee-library-harvard-google-smithsonian-digitize/>

collections content, although one library noted that they sense from conversations with HathiTrust staff that this is changing.

- Cost was cited as a major concern for libraries in this broad category. These libraries will be charged in the new cost model based on their overall print holdings overlap with HathiTrust. For smaller libraries, this may not sum to a huge fee, but these library interviewees noted a break in the linkage between their perception of the service they would get from HathiTrust (hosting and preservation for a subset of special collections material) and the price they would be asked to pay (which will be linked to overall print holdings). It is important to note that these interviews were conducted before the universities of Michigan, Wisconsin, and Florida announced their intention to begin making orphan works available to their authorized campus users via HathiTrust; libraries in this category may well judge the value of HathiTrust differently now that there may be additional access benefits to HathiTrust beyond the existing preservation and other benefits.

*Libraries that participate in regional or consortial digital library projects, and see HathiTrust as duplicative.* These libraries may have robust digitization programs, including some scanning of general collections, but they receive repository services from a consortial or regional digital library. The fees for these solutions may be wrapped into the participation fees for the consortium, or, as in at least two cases with our interviewees, the costs may be paid by a governmental body.

- Those libraries active in consortia voiced some hesitance to make a move without the consent and participation of their partners in existing arrangements: “We’ve invested heavily in our consortium, both financially and philosophically.” There is a flip side to this, in that two such interviewees also noted that they would expect to receive a consortial discount on HathiTrust participation fees.
- The cost of participation in HathiTrust is also a concern for this group, particularly for those that receive free or low-cost digital storage through a governmental body. “For all the emotional reasons, I would like to support [HathiTrust],” one interviewee said. “But what do I show my faculty and staff we’re getting for the money?” As the process for making orphan works accessible in HathiTrust becomes clearer, and as the Universities of Indiana and Illinois launch the HathiTrust Research Center, designed to facilitate text mining across the corpus, libraries will be able to more clearly weigh the access benefits when making a decision about membership.
- Yet, these same libraries noted some interest in using the availability of digitized copies in HathiTrust as a print collections management tool, and a general interest in supporting a community-driven initiative. (One interviewee stated this in opposite terms: “We don’t want to be seen as a free-rider.”)

*Libraries that do not have a strong preservation emphasis.* These libraries simply may not have a strong interest in devoting funds to an initiative around legacy print collections. As an interviewee at one such library put it: “Our priorities are on supporting emerging modes of scholarly communication,” with the implicit belief that HathiTrust is primarily a preservation solution for legacy materials.

- These libraries may have an interest in new initiatives like the Research Center and, perhaps, in the ability to open up access to orphan work..

Several concerns cut across these categories:

- *Confusion about the proposed cost model.* In particular, library representatives wondered what would happen if other libraries were to exit HathiTrust—which they see as a distinct possibility in what remains a difficult economic climate for academic libraries. The number of partner libraries is an inverse driver of costs in the proposed model—that is, in general, the more libraries, the lower the cost for each individual library—but, as several interviewees pointed out, the model does not include a control for the number of libraries used to calculate HathiTrust membership fees. In general, clearer explanation and a great deal of outreach to libraries who are concerned about this would be needed in order to convince them to join.
- *For libraries outside the United States, a concern about joining a U.S.-based initiative.* One library felt that they were precluded from joining HathiTrust by European law; a Canadian research library noted a general hesitancy among their peers to place their content on U.S.-based servers, given the different cultural and legal context. Intensified outreach efforts would be needed to allay concerns like these if the HathiTrust leadership decides that they would like to greatly increase international participation in the near term.

### Other Strategic Concerns

Partner libraries clearly value the preservation function of HathiTrust—in the survey of participating libraries, this was listed most often as a library’s principal motivation for joining HathiTrust.<sup>11</sup> But access to the works was also clearly stated as a value by the library representatives who participated in our research, and this was particularly true for libraries that do not yet participate: several of the libraries do not have a strong research and preservation focus, but want to be able to make a wider volume of the content they own in print available to their users in digital form.

During the course of this research, the universities of Michigan, Wisconsin, and Florida, announced that they would make digitized copies of orphan works available to authorized users via HathiTrust, a move that could theoretically be adopted by other HathiTrust partner libraries, as well.<sup>12</sup> If this capacity can be extended to other libraries, then HathiTrust can clearly address the needs of the larger constituency of libraries that do not have a great deal of their own digitized content to contribute. Indeed, this offers the HathiTrust staff an opportunity to use print holdings comparisons analysis (using WorldCat or other means) to identify a priority set of libraries to pursue for the HathiTrust partnership: those libraries that have local collections with a large proportion of orphan works and that would greatly benefit from a HathiTrust initiative to help libraries open up access to orphan works for which the partner library owns a copy.

---

<sup>11</sup> See Appendix C, questions 3 and 4.

<sup>12</sup> <http://blogs.uflib.ufl.edu/news/2011/07/14/university-of-florida-libraries-join-the-hathitrust-digital-library-and-expand-digital-access-to-%E2%80%9Corphan%E2%80%9D-works/>

Given the strong statements by interviewees of the potential value of access to digitized works held in HathiTrust, the HathiTrust staff and community may wish to focus their limited outreach time on the libraries that would be identified by this analysis.

In addition to the opportunities provided by the orphan works access initiative, there are at least two other avenues that HathiTrust could consider in order to build on the value of the initiative as a digital preservation solution:

- *More actively pursuing permissions from publishers.* HathiTrust has already experimented with this in a sense, as some books from the University of Michigan Press (itself tightly integrated with the Michigan Libraries) are made openly available in HathiTrust. While establishing a Publisher Relations function within HathiTrust would be a significant shift for the project, pursuing agreements with publishers would allow HathiTrust to open a greater amount of content to end-users while strengthening HathiTrust's value proposition as a preservation service to its library constituents.
- *Focusing on ingesting special collections, possibly including formats other than books.* Libraries are increasingly digitizing special collections content, and it may be that HathiTrust has a role to play in serving as both a preservation and access aggregator for this content. (Indeed, several interviewees from libraries that do not yet participate in HathiTrust wondered aloud whether HathiTrust could accept their special collections content.) There are several contingencies to such a plan. First, several library representatives noted the time-cost of preparing files for ingest into HathiTrust as a concern, so partner libraries would need to be apprised of, and comfortable with, that time burden. In addition, there are questions about how to effectively present special collections content to facilitate use in research and teaching (particularly with multiple format types if that is contemplated). In Ithaka S+R's past client research for primary source collections, we have conducted extensive background research on existing primary source products for teaching and scholarship, and on faculty preferences for such collections. The overarching concern for such products is that, in order to be effective for a broader range of end-users than researchers, there needs to be some form of editorial shaping, in the form of curated collections of content, suggested excerpts for classroom or student assignment use, and/or contextual explanatory material sitting around the primary sources. Merely aggregating a collection of special collections or primary source materials is not likely to provide a coherent experience for the end-user or drive the level of usage that would demonstrate to partner libraries the impact and value of placing their collections in HathiTrust.
- *Developing a comprehensive strategy for tangible materials.* There is already work underway to determine if HathiTrust can provide for the preservation of print originals following their digitization. Ithaka S+R has seen some indications that supporting print preservation via business models organized principally around access to and preservation of digitized versions may be sustainable (similar to approaches taken by LLMC and JSTOR). In addition, based on the responses of library directors to the Ithaka S+R Library Survey, HathiTrust may find it valuable to offer opportunities to provide for on-demand access to

these print originals, at least to HathiTrust members, whether directly or in partnership with print sharing networks.

### **Potential Next Steps and Questions for the HathiTrust Community**

The purpose of this report is not to prescribe action for HathiTrust—“our survey and interviews seem to say *x*, therefore HathiTrust should do *y*”—but rather to surface opportunities and challenges, and offer those to the HathiTrust community in a way that allows the Strategic Advisory Board and the working groups to shape the agenda for the Constitutional Convention. With that proviso, our work suggests that the following items may be helpful for the HathiTrust community to consider. It also bears noting that each of these is interrelated: that is, questions about mission are closely tied to questions about the governance model, the size and composition of the membership, and the sustainability model that will be needed to support the archive.

- *Clearly defining objectives for the next 3-5 years, possibly mapping out the rationale for those objectives in the context of a revised mission statement.* In short, partner libraries want to know where HathiTrust is headed. All of the other questions or next steps that we suggest flow from this fundamental one: what will HathiTrust be in the future? Which of its six goals (quoted on the first page of this briefing paper) will prove to be the ones that drive the collaboration? While these six items are not necessarily mutually exclusive, it is easy to envision HathiTrust taking on initiatives that would align with HathiTrust’s mission goals as stated now—for example, a coordinating function for shared print storage—that might cause current and potential partner libraries to reassess HathiTrust’s value and costs.
- *Enhancing the pipeline of information about HathiTrust’s strategic priorities to partner libraries.* Respondents to the survey and follow-up interviewees spoke in particular about the need for clear communication from the HathiTrust staff and leadership, noting in several cases that they have not always felt well apprised of contemplated new initiatives. One interviewee said that their institution is quite eager to commit staff time to a HathiTrust development initiative, but they do not fully understand the menu and timeline of contemplated initiatives, and how they might effectively “plug in” their developers. Another interviewee tied a perceived lack of communication about new initiatives to a larger question about mission creep (e.g., “Will HathiTrust be taking on new content types? And if so, why?”). To the extent that HathiTrust is able to provide a more constant flow of information about proposed initiatives, that would seem to be valued by partner libraries.
- *If HathiTrust is to retain a membership-driven governance structure, there might be value in advancing changes that allow for nomination and selection to the Executive Committee, and clearly communicating the options to the HathiTrust partner community ahead of the Constitutional Convention.* A number of libraries, particularly those that have contributed content to the archive, noted that they expect heavy content contributors and/or founding members to have a more heavily weighted voice in any future governance model for HathiTrust.

- *Demonstrating to partner libraries the sustainability and feasibility of the new cost model for HathiTrust.* Some current participants raised concern about the potential for increased membership fees over time, and libraries that do not yet participate worried about a perceived instability in the pricing structure. If HathiTrust elects to move forward with a version of the proposed cost model, there may be value in introducing a stabilizer into the cost model—either asking that institutions commit to membership for a set period, or calculating the divisor in the formula for public domain volumes (the number of institutions participating in HathiTrust) using a rolling average that takes into account anticipated uptake in the next year.
- *Ultimately, making a decision based on the most pressing goals and objectives for HathiTrust, about how large the membership for the initiative needs to grow.* The proposed transitional cost model privileges the addition of new members, as each additional member helps to lower the costs for other libraries—an elegant example of network effects benefitting the HathiTrust membership. On the other hand, adding new institutions may introduce new challenges for HathiTrust, particularly if:
  - a wider range of content beyond general collections print holdings is taken into the archive, potentially increasing the overall costs for member libraries even if they only value HathiTrust for preservation of books, serials, or government documents, or;
  - if HathiTrust retains a membership-driven governance structure *and* expands membership beyond the research library community, in which case the need to accommodate libraries with modest budgets and a desire to have a voice in leadership may be at odds with the transitional cost model and a challenge to a governance structure that may privilege the voice of heavy content-contributing research libraries.

We raise these questions in order to prompt conversations around the future shape of the initiative and the role that current and future partner libraries will play in governing and sustaining HathiTrust. We thank all of the interviewees and survey respondents for their input throughout the project.



## Appendix A. Interviewees

### HathiTrust Participating Libraries

Bob Wolven, Associate University Librarian, and Janet Gertz, Director of Preservation and Digital Conversion Program, Columbia University (G,M)  
 Carol Hughes, Associate University Librarian, University of California-Irvine (G,M)  
 Heather Christenson, Mass Digitization Project Manager, California Digital Library (G,M)  
 John Wilkin, Executive Director of Hathi Trust and Associate University Librarian, and Jeremy York, Project Librarian for HathiTrust, University of Michigan (G)  
 Lisa German, Assistant Dean for Technical Services and Collections, Pennsylvania State University (G)  
 Rick Luce, Vice Provost and Director of Libraries, and Lars Meyer, Senior Director for Content, Emory University  
 Ruth Jackson, University Librarian, University of California-Riverside (G,M)  
 Thomas Leonard, University Librarian, University of California-Berkeley (G,M)

### Other Libraries

Artemis Kirk, University Librarian, Georgetown University  
 Carla Stoffle, Dean of Libraries and Center for Creative Photography, and Stephen Bosch, Budget and Licensing Librarian, University of Arizona  
 Cathrine Harboe-Ree, University Librarian, Monash University, Australia  
 Christine A. Fowler, Digitisation Manager, University of Southampton, U.K.  
 Harriette Hemmasi, University Librarian, and Andrew Ashton, Director of Digital Technologies, Brown University  
 Jeannette Frey, University Librarian, Lausanne Cantonal and University Library (G)  
 Jeff Hiroshi Gima, Director, American International Consortium of Academic Libraries<sup>13</sup>  
 Jonathan LeBreton, Senior Associate University Librarian, Temple University  
 Judith Russell, Dean of Libraries, University of Florida  
 Mary-Jo Romaniuk, Acting Chief Librarian, University of Alberta, Canada (M)  
 Patricia Killiard, Head of Electronic Services and Systems, and Grant Young, Digitisation and Digital Preservation Specialist, University of Cambridge, U.K.  
 Patricia Tully, University Librarian, Wesleyan University  
 Robert Kieft, College Librarian, Occidental College  
 Sabrina Pape, College Librarian, Vassar College

### Researchers/Subject Librarians

Fred Burchsted, Reference Librarian and Subject Specialist for History, Harvard University  
 Katherine Rowe, Professor of English, Bryn Mawr College  
 Kristin Eschenfelder, Assistant Professor of Information, University of Wisconsin-Madison  
 Joseph Janes, Associate Professor of Information, University of Washington  
 Miles Efron, Assistant Professor of Information and Library Sciences, University of Illinois, Urbana-Champaign

(G) denotes a library that participated in the Google Books project.

(M) denotes a library that participated in the Microsoft Live Search digitization project. (Alberta partnered with the University of Toronto on its Microsoft digitization project.)

---

<sup>13</sup> The American International Consortium of Academic Libraries, or AMICAL, is a group of teaching-focused higher education institutions in Europe and the Middle East.

## **Appendix B. *Alternative organizational models***

The Strategic Advisory Board requested that Ithaka S+R provide a brief overview of potential alternative governance models for an organization like HathiTrust. (To prescribe one of these models—or even to argue definitively that the current model is insufficient for HathiTrust’s future needs—is outside the scope of this project.) Below, we suggest three broad organizational models that the Board may wish to consider, arranged roughly in order from least centralized to most centralized.

*A consortium.* In a typical consortium, a group of institutions pools some level of resources to work toward a set of shared goals, often with some amount of central staffing to support the programmatic initiatives of its members. Consortia with deep programmatic emphases typically serve a fairly fixed number of members working together to pursue shared objectives in a trusted environment. Examples of strong consortia with programmatic emphases include CDL and CIC, both of which are governed at least to some degree at the institutional level in addition to the library level. Other consortia may be governed with equal votes of all the member libraries, or via other mechanisms.

*A membership organization.* Whereas some consortia are governed fairly informally and may not have independent legal status, a membership organization is a formally incorporated not-for-profit. In a membership organization, the constituents who make up the collective have varying degrees of decision-making power and channels to provide input, but they nearly always have the right both to approve bylaws to the organization’s constitution, and to elect members to the organization’s board of trustees. One of the largest and best-known such organizations in the library world is OCLC, which is overseen by a governing board, the members of which are elected by the OCLC membership. (OCLC now counts more than 25,000 institutions in its membership.<sup>14</sup>) At larger membership organizations, there may be nominating committees to help in the election process to the governing board.

*An independent not-for-profit organization.* Like a membership organization, the independent not-for-profit organization is ultimately governed by a board of trustees; in this case, though, that board of trustees is self-perpetuating, electing its own members as vacancies arise. One trait of self-perpetuation is that it can help to provide for somewhat more strategic continuity than certain types of membership organizations. At the same time, independent not-for-profit organizations are not governed directly by the participants and clients they seek to serve.

---

<sup>14</sup> <http://www.oclc.org/us/en/about/cooperation/default.htm>