



# HATHI TRUST

## HATHITRUST COLLECTIONS COMMITTEE

### Collection Priorities Survey Analysis

---

May 11, 2016

*Carmelita Pickett, University of Iowa, Chair*  
*Martha Hruska, University of California, San Diego, PSC Liaison*  
*Sharon Farb, University of California, Los Angeles*  
*Bryan Skib, University of Michigan*  
*Claire Stewart, University of Minnesota*  
*Tom Teper, University of Illinois at Urbana-Champaign*

## HathiTrust Collections Survey Analysis

### Introduction

The HathiTrust Program Steering Committee (PSC) charged the Collections Committee with developing a survey for members regarding current and prospective collection priorities. Following significant discussion between the Collections Committee and PSC in order to define the survey's scope the finished survey was sent to all members on October 6, 2015. The survey closed on November 6, 2015, with a total of seventy-six (76) full or partial responses received from 136 members (a response rate of 55.9%). The survey's design sought to intentionally gather both quantitative and qualitative responses, with open ended qualitative responses that would elicit detailed comments not included in quantitative responses. The report contains an appendix with visualizations of the qualitative questions and responses. In addition to the appendix, this document identifies common findings in the responses, discusses the respondents' notions of core business vs. future business, highlights responses related to quality issues with HathiTrust content, and makes several recommendations based on the responses received.

### Top Findings

Based upon the responses received, the top six trends reflected in the survey are the following:

1. **Current Collection Strategy** - HathiTrust members overwhelmingly (93%) support the current collection strategy and scope. HathiTrust's focus on collecting and preserving published book and journal type content with the intent of support for teaching and research while providing the broadest level of public access possible remains highly valued among the membership.
2. **Gap Filling** - Related to this top trend are comments from members expressing interest and support for HathiTrust to concentrate efforts on filling in the gaps with respect to the corpus of published works. Such gaps include missing volumes from sets, oversized materials, missing fold-outs, and items rejected for scanning based on condition-related issues (aka Google or Internet Archive (IA) "rejects" from standard scanning processes).
3. **Quality** - HathiTrust members expressed strong support for enhancing the quality of the existing digitized titles (e.g. missing volumes, pages, etc.).
4. **Documentation** - HathiTrust members expressed strong interest in and some frustration at the lack of easy and well established policies, procedure and practices to provide easy uploading to HathiTrust of member generated scans that are outside of the established Google and IA workflows.
5. **Accessibility** - HathiTrust members voiced similar concerns over the perceived lack of sufficient documentation and user guides as posing barriers to access for people with print disabilities.
6. **Expanded Mission and Cost** - HathiTrust members expressed concern regarding the potential costs of HathiTrust expanding collection development priorities for such as multimedia, images, manuscripts etc.

## **Core Business vs Future Business**

HathiTrust has built a strong brand around what is – primarily – a mechanism for collectively storing and delivering digitized print materials. The success of this brand has encouraged members to consider opportunities for expansion into other services around storing, discovering and delivering non-book digital collections. As such, the survey included questions aimed at priming members to consider HathiTrust’s current collection focus and potential support for new collection priorities. Based on survey results 93% (Q9) of respondents agree that HathiTrust should continue to expand its current digitized print-based corpus.

The survey asked members to consider the addition of new content to expand the HathiTrust corpus. Two clear priorities emerged over other options presented. The members expressed a desire to prioritize mass digitization “rejects” (e.g., materials rejected due to condition or size) followed by targeting specific corpora recommended by scholars, over other options. Members seemed to express mixed interest about open access books, born digital monographs, and in-copyright books as a strategy for expanding the current corpus.

HathiTrust members were asked to express their interest in using the repository to preserve the following content (Q11): web archives, executable content, encoded texts, maps, manuscript and archival materials, ephemera, still images, moving images, and audio materials. Members seem to value manuscript and archival materials as the leading priority followed by maps, still images, and moving images. A similar pattern emerges in member responses to question 12. Members are keenly interested in HathiTrust’s ability to provide access to manuscript and archival collections. There seems to be less interest in executable content and web archives. Members support HathiTrust as a collection solution to address materials at risk for preservation and see the repository as a benefit to end users of aggregated access (Q 13). When considering new content types, members view HathiTrust as a logical collective solution. This held true regardless of whether local infrastructures were inadequate or too costly to maintain, or in instances when collective solutions were just viewed as more cost effective mechanisms for addressing challenges they hold in common.

Beyond current investments in digitized print-based corpus, some members would like to see HathiTrust invest in collection development tools, collaborative collection building, and the storage of high-resolution TIFFs. These efforts appear to have a higher value for members (Q 15) than many other options provided. That said, members also indicated interest in limited investment in open access content development, eliminating duplicates, and ingesting library published content.

It is clear that members view HathiTrust as the primary basis for developing a collective collection of print monographs (Q15). Members expressed a desire to leverage HathiTrust to inform local decisions, although the path toward that was not specified in the responses.

If new business initiatives are explored and implemented by HathiTrust, members caution that cost to the members should factor into the decision-making process.

### **Quality: Metadata and images**

While generally pleased with and committed to expansion of the current book corpus, HathiTrust members did voice concerns about the quality or suitability of the content – and in particular the metadata. Responses to questions 10 (a) show priority interest in filling gaps in multi-volume sets (and serials), followed by correction of page-level errors, and then the reinsertion of missing foldouts. The survey’s structure required assignment of a different level of interest to each quality concern, even though some institutions would have assigned a high priority to multiple areas suggested for quality improvement. One partner, for example, noted in later free-text comments that foldouts, missing pages and gaps in multi-volume sets were all very important areas for HathiTrust’s attention. Another response indicated that their highest priority was for stronger quality control and assurance. Free-text responses to later survey questions noted the “foldout problem,” and several partners recommended improvements in the metadata for serials and multi-volume sets, with a more logical and standardized ordering seen as a boon for user discovery and for collection management decisions. Completeness of individual digital facsimiles was further noted as an important consideration when considering withdrawal of print copies. One partner asked that HathiTrust focus on “better access” within the current scope, prior to expanding that scope. Several responses to the final survey question (Q17) voiced generalized concern about the completeness and quality of both content and metadata. Two partners suggested the enhancement of metadata for rare books, primarily to support disambiguation, the inclusion of copy-specific details, and additional searchable fields should also be considered in an effort to improve quality issues.

### **Recommendations**

Given HathiTrust’s stated vision of providing a ‘comprehensive archive of published literature from around the world’ that can support ‘shared strategies for managing and developing the digital and print holdings in a collaborative way’, members of the Collections Committee believe that the survey results indicate a clear need for the organization to be more strategic in further developing this archive. To that end, we recommend that PSC consider the following recommendations:

***Concentrate on Enhancing the Comprehensiveness of the Digitized Print Corpus*** – The comprehensiveness of the print corpus continues to be a significant point of concern for the membership. This includes global expansions of the corpus such as providing mechanisms for the membership to incorporate materials such as manuscripts, archival records, and scholarly works such as theses and dissertations. It also includes seeking to identify “spotty” areas in the corpus that need attention. Particular subjects, languages, and regions may be lacking, and scholars could help scope how extensive the printed corpus might be in these categories relative to what exists within Hathi. In all cases, improvements to these categories would be facilitated by improving the processes by which members can contribute content.

***Improving the Quality of the Corpus*** – Significant concerns remain regarding the quality of the print corpus and the metadata supporting discovery. These concerns range from missing items (from sets) to missing images, foldouts, and pages. The ingestion process for correcting these deficiencies remains lacking. Additionally, the quality of the metadata that underlies many of our items is viewed as lacking as well.

***Improving Member Services*** – Improvements to the items listed above will generate ancillary benefits to the members and provide greater opportunities to enhance member services. One aspect of member service that emerged from the survey responses was a greater desire to see services for those with print disabilities improved. Currently, the mechanism is perceived as cumbersome. Additionally, the membership believes that opportunities exist for HathiTrust to develop member services such as collection development analysis tools that would provide better services than those currently on the market.

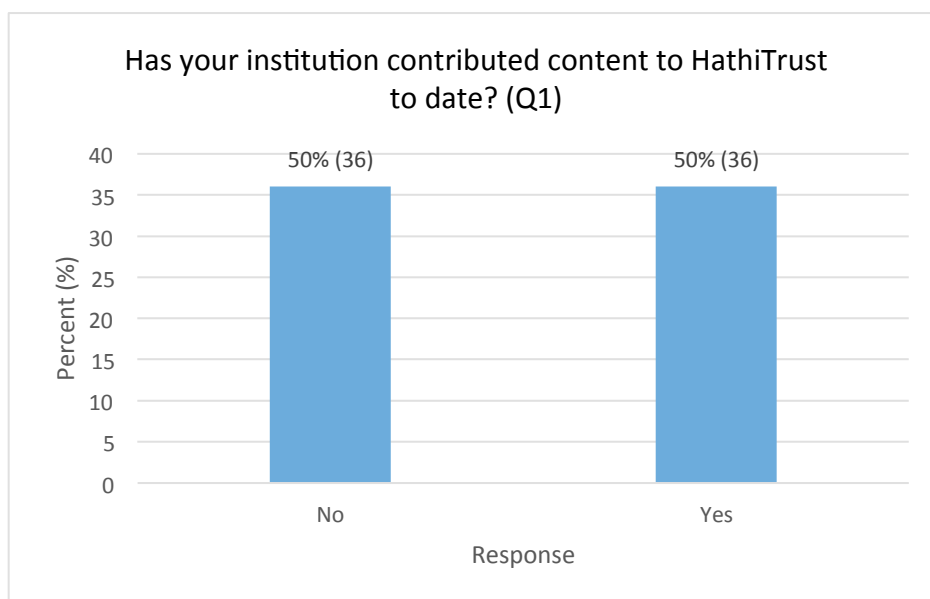
### **Supplemental Notes**

Based on the Collections Committee's experience in compiling the results of this survey, this committee recommends that HathiTrust plan to seek professional the expertise of those who develop, run, analyze and visualize complex survey data for large member based organizations when future surveys of this nature are contemplated. The Collections Committee is expert on the subject matter related to the survey. However, as this work progressed, it was apparent to our Committee that we did not have the survey expertise necessary for a detailed analysis of all the free text, nor the experience with how best to present or visualize data of this scope and complexity.

## Appendix A: Charts & Graphs

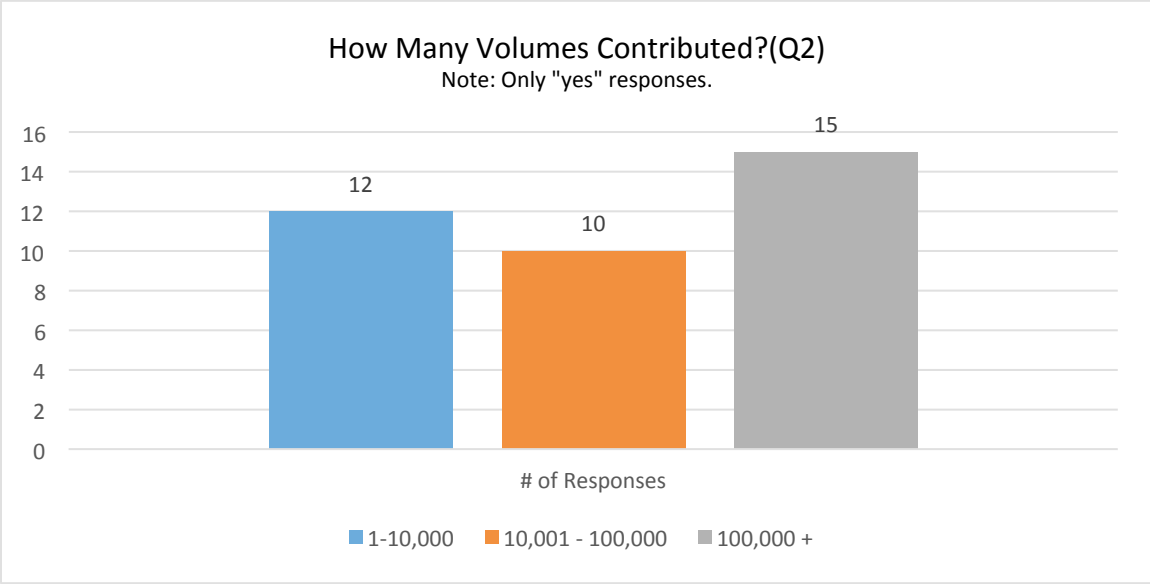
### Question 1

<b>Has your institution contributed content to HathiTrust to date? (Q1)</b>				
No	Percent	Yes	Percent	Total
36	50%	36	50%	72



### Question 2

<b>How Many Volumes Contributed? (Q2)*</b>		
Vols	# of Responses	Percent
1-10,000	12	32.00%
10,001 - 100,000	10	27.00%
100,000 +	15	41.00%
<b>Total</b>	<b>37</b>	<b>100.00%</b>
*Note: Only yes responses.		

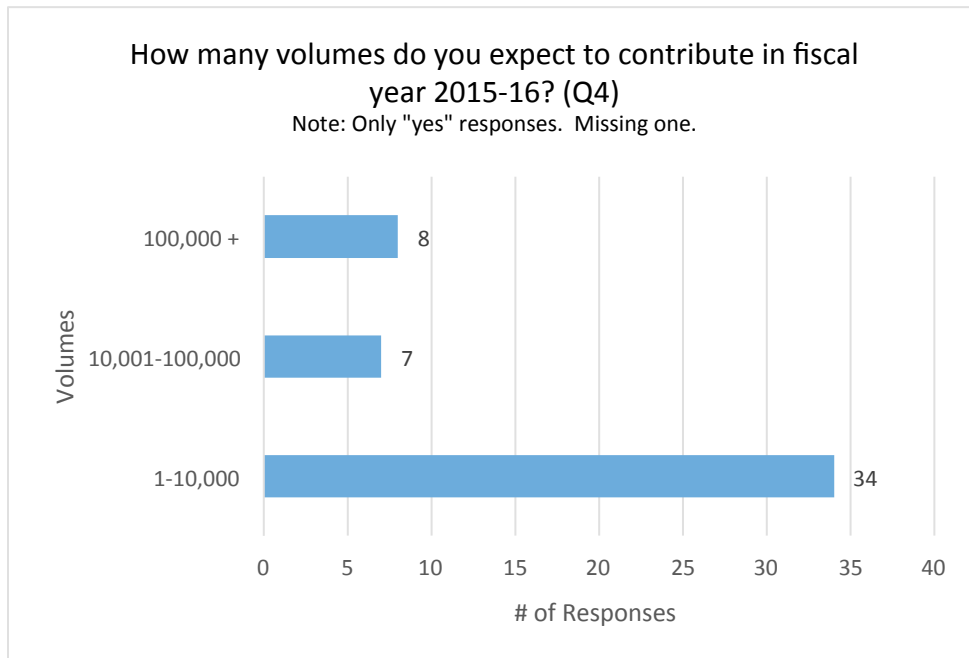


**Question 3**

<b>Do you anticipate contributing volumes to HathiTrust in the next year? (Q3)*</b>		
Response	Number	Percent
Yes	48	67.00%
No	23	32.00%
Total	71	99.00%
*Note: Missing one response		

**Question 4**

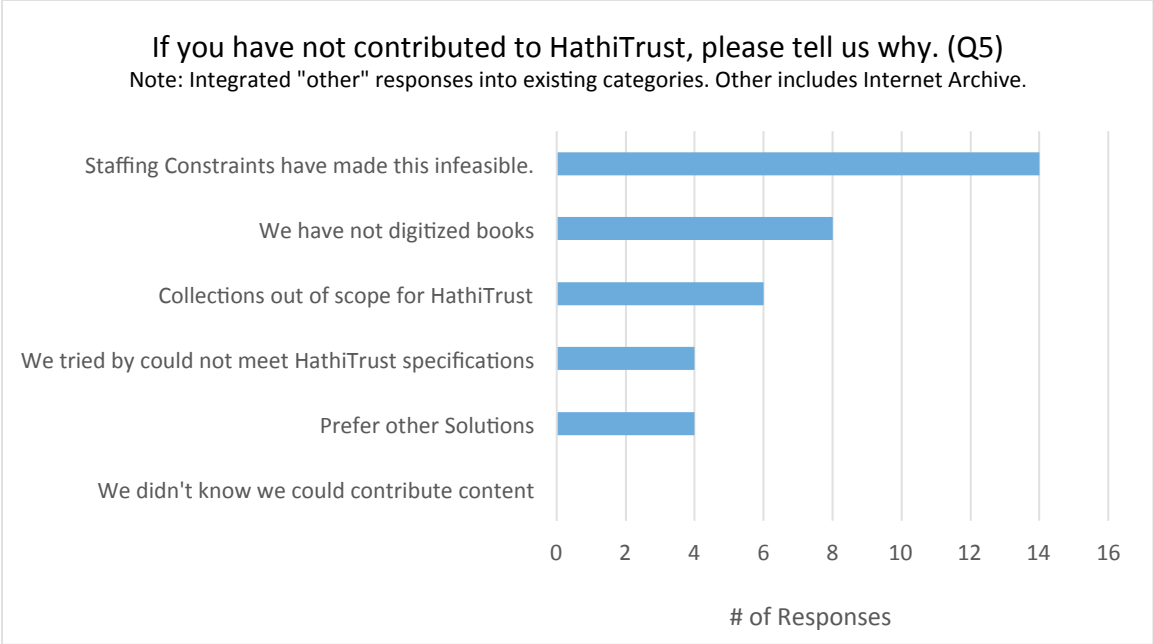
<b>How many volumes do you expect to contribute in fiscal year 2015-16? (Q4)</b>		
Vols	# of Responses	Percent
1-10,000	34	69.40%
10,001-100,000	7	14.28%
100,000 +	8	16.32%
Total	49	100.00%



**Question 5**

<b>If you have not contributed to HathiTrust, please tell us why. (Q5)*</b>	
Responses	# of Responses
We didn't know we could contribute content	0
Prefer other Solutions	4
We tried by could not meet HathiTrust specifications	4
Collections out of scope for HathiTrust	6
We have not digitized books	8
Staffing Constraints have made this infeasible.	14
<b>Total</b>	<b>36</b>
Note: Integrated other responses into existing categories. Only 36 responses. Other includes Internet Archive	

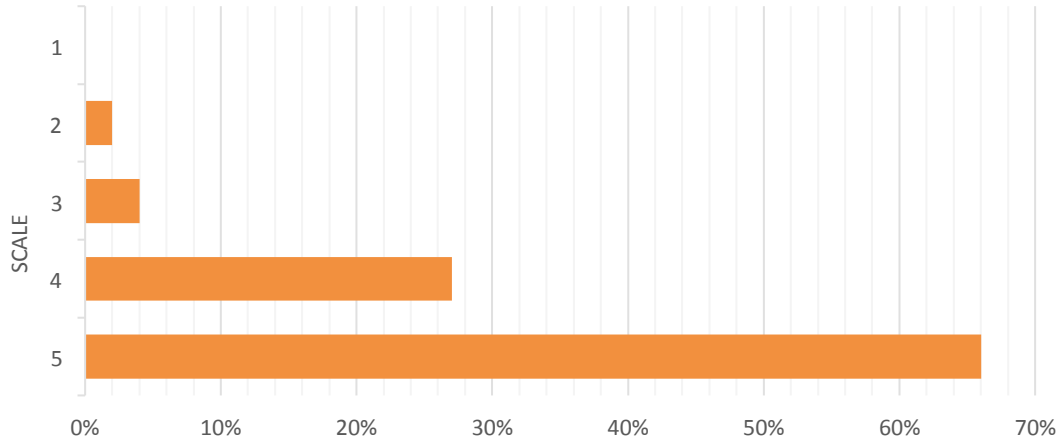




**Question 9**

<b>How important is it to your institution that HathiTrust continue to develop and expand the current print-based book corpus? (Q 9)</b>		
Scale (1=Not at all important; 5=Very important)	Responses	Percent
5	49	66%
4	20	27%
3	3	4%
2	2	2%
1	0	0%

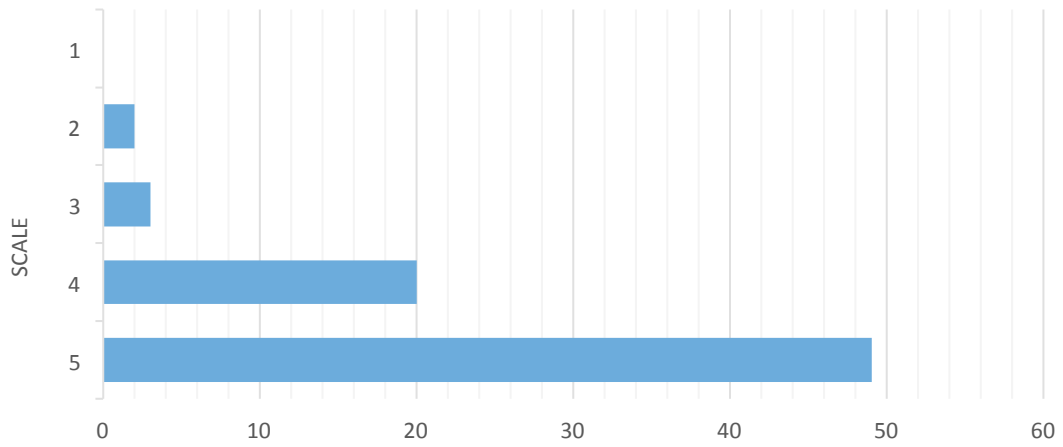
How important is it to your institution that HathiTrust continue to develop and expand the current print-based book corpus? (Q9)  
 1=Not at all Important; 5=Very Important



	5	4	3	2	1
■ Percent	66%	27%	4%	2%	0%

RESPONSE

How important is it to your institution that HathiTrust continue to develop and expand the current print-based book corpus? (Q9)  
 1=Not at all Important; 5=Very Important

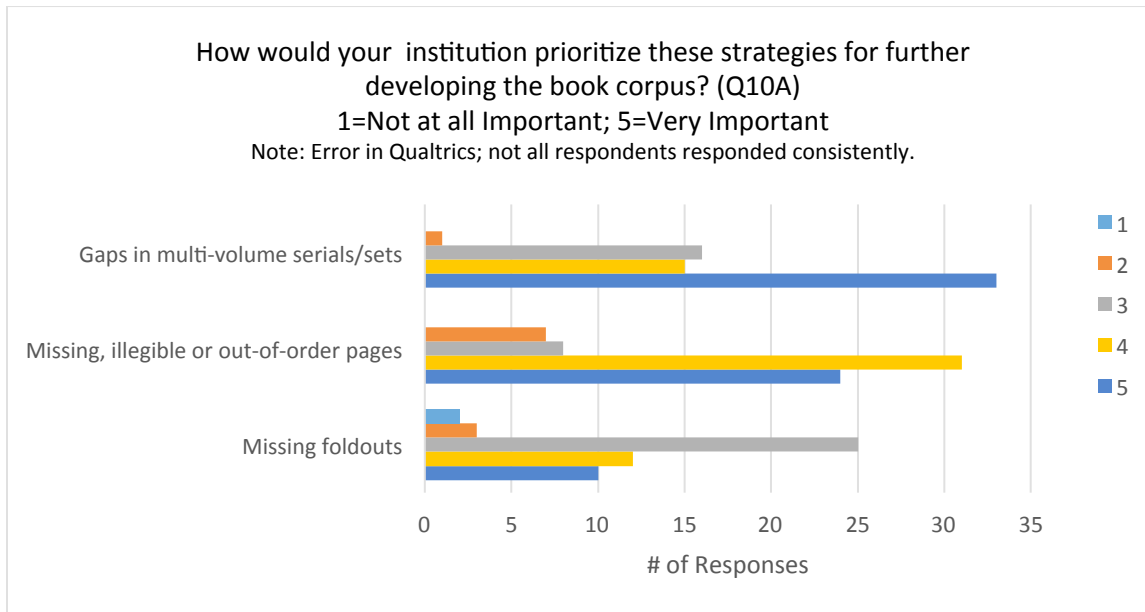


	5	4	3	2	1
■ Response	49	20	3	2	0

RESPONSE

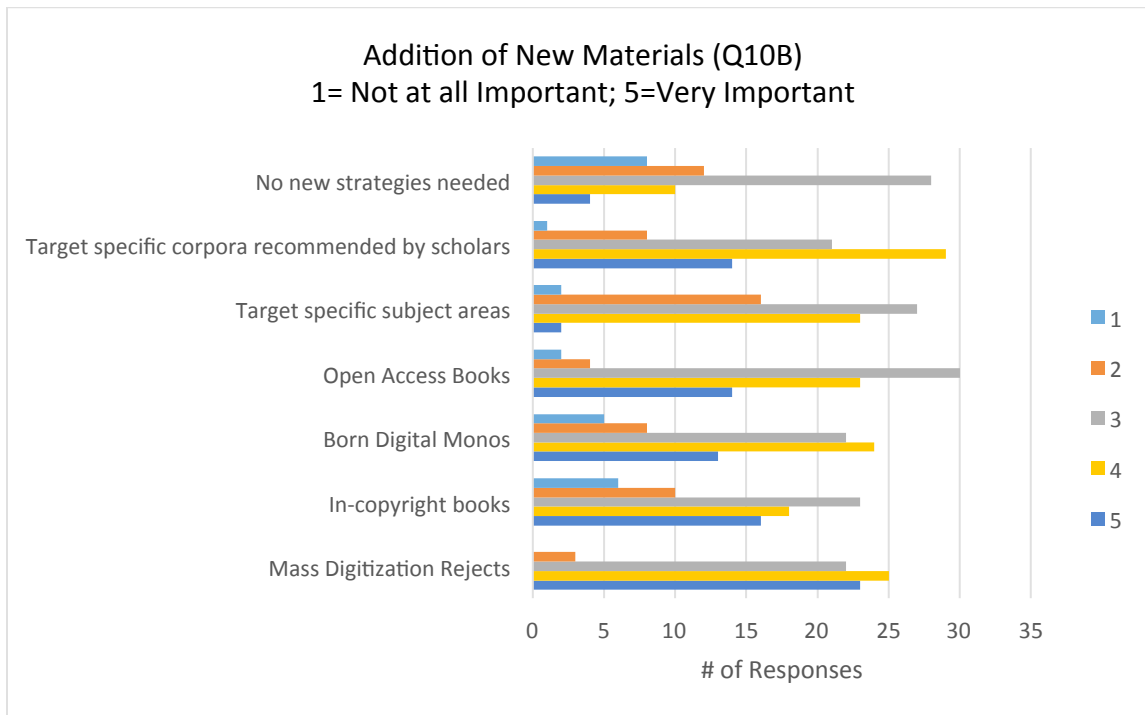
### Question 10A

How would your institution prioritize the following strategies for further developing the book corpus? Scale 1 to 5 (1=Not Important; 5=Very Important) Quality Improvement of the existing corpus. (Q 10 A)*			
	Missing foldouts	Missing, illegible or out-of-order pages	Gaps in multi-volume serials/sets
1	2	0	0
2	3	7	1
3	25	8	16
4	12	31	15
5	10	24	33
Total	52	70	65
Avg. Response	3.48	4.03	4.23
Note: Error in Qualtrics; not all respondents responded consistently.			



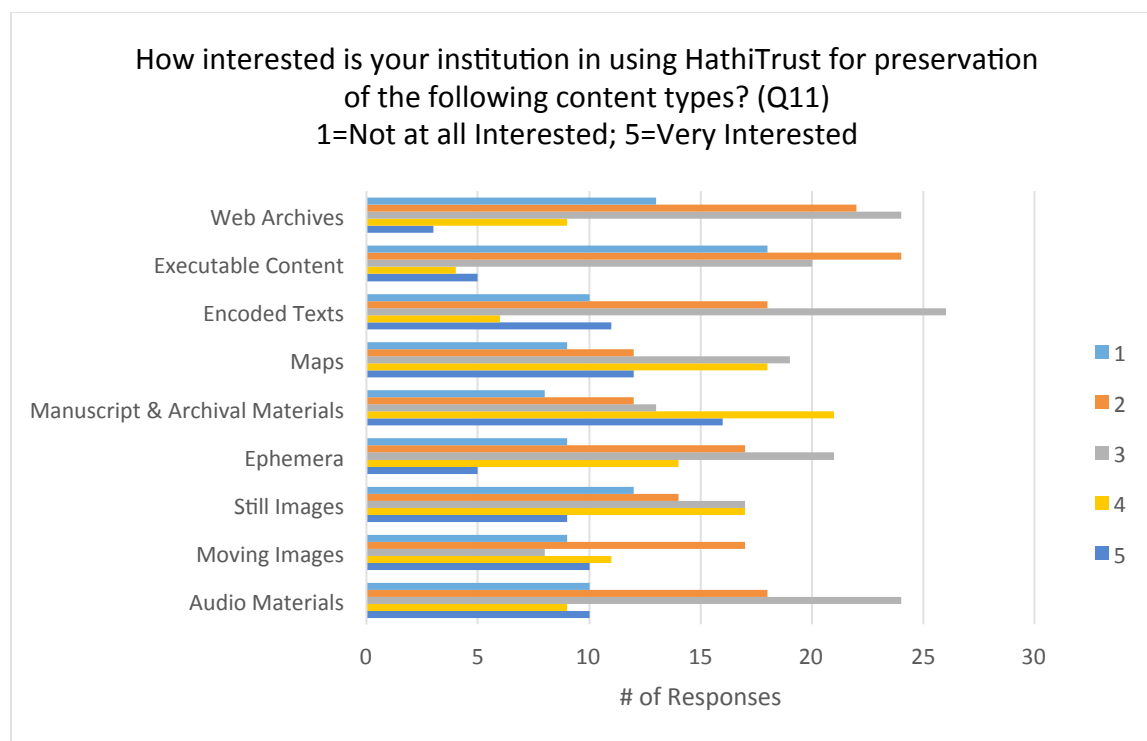
### Question 10B

How would your institution prioritize the following strategies for further developing the book corpus? Scale 1 to 5 Addition of new materials? (Q10B)							
	Mass Digitization Rejects	In-copyright books	Born Digital Monos	Open Access Books	Target specific subject areas	Target specific corpora recommended by scholars	No new strategies needed
1	0	6	5	2	2	1	8
2	3	10	8	4	16	8	12
3	22	23	22	30	27	21	28
4	25	18	24	23	23	29	10
5	23	16	13	14	2	14	4
Total	73	73	72	73	70	73	62
Avg.	3.93	3.38	3.44	3.59	3.1	3.64	2.84



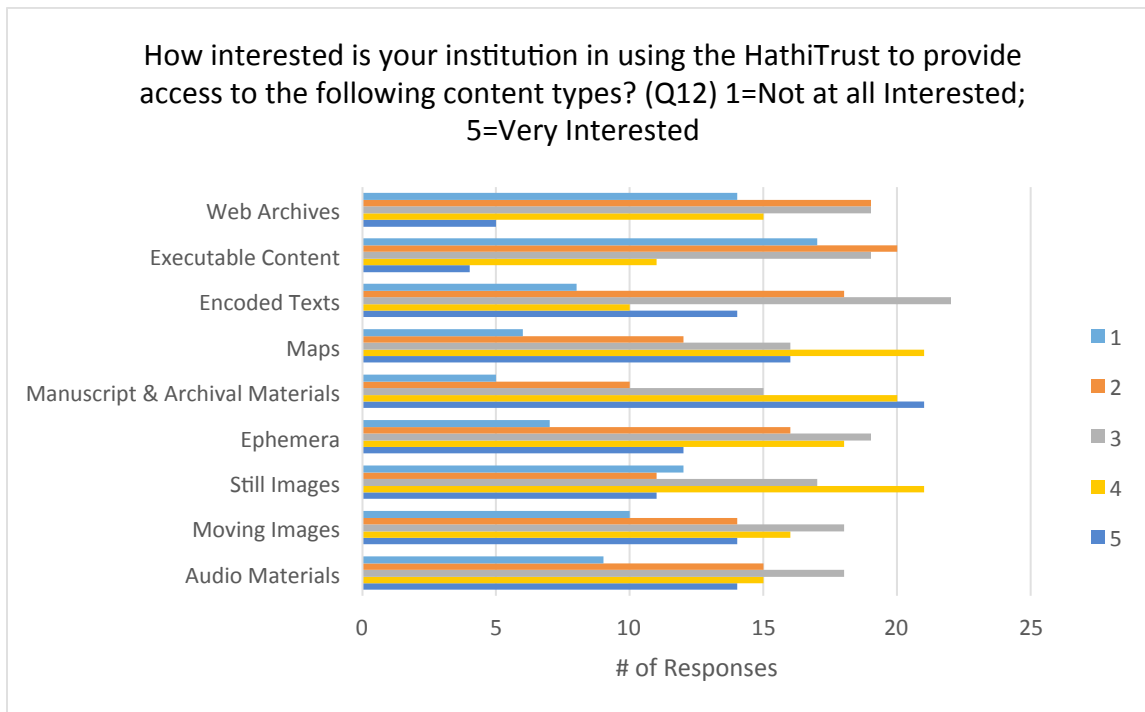
### Question 11

How interested is your institution in using HathiTrust for preservation of the following content types?							
Scale 1=Not at all Interested; 5=Very Interested (Q11)							
	1	2	3	4	5	Total	Avg. Resp.
Audio Materials	10	18	24	9	10	71	2.87
Moving Images	9	17	8	11	10	55	2.93
Still Images	12	14	17	17	9	69	2.96
Ephemera	9	17	21	14	5	66	2.83
Manuscript & Archival Materials	8	12	13	21	16	70	3.36
Maps	9	12	19	18	12	70	3.17
Encoded Texts	10	18	26	6	11	71	2.86
Executable Content	18	24	20	4	5	71	2.35
Web Archives	13	22	24	9	3	71	2.54



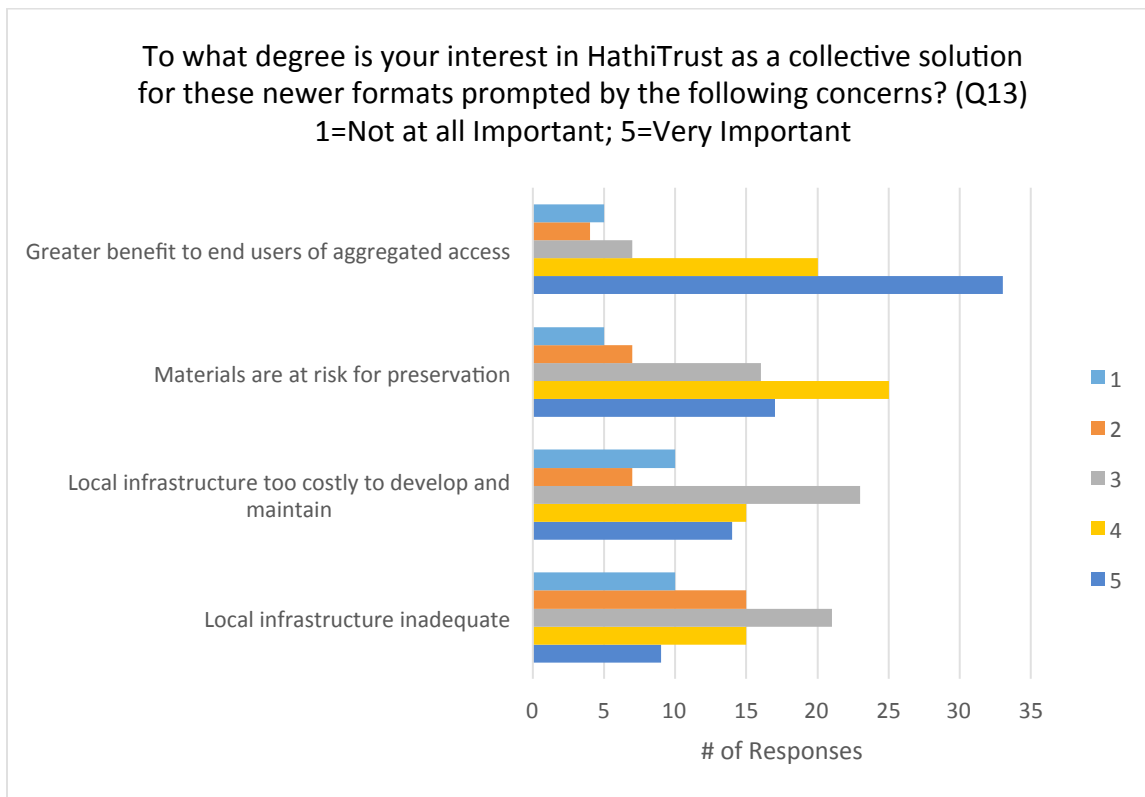
### Question 12

How interested is your institution in using the Hathitrust to provide access to the following content types? (Q12) Scale 1= Not at all Interested; 5=Very Interested (Q12)							
	1	2	3	4	5	Total	Avg. Resp.
Audio Materials	9	15	18	15	14	71	3.14
Moving Images	10	14	18	16	14	72	3.14
Still Images	12	11	17	21	11	72	3.11
Ephemera	7	16	19	18	12	72	3.17
Manuscript & Archival Materials	5	10	15	20	21	71	3.59
Maps	6	12	16	21	16	71	3.41
Encoded Texts	8	18	22	10	14	72	3.06
Executable Content	17	20	19	11	4	71	2.51
Web Archives	14	19	19	15	5	72	2.69

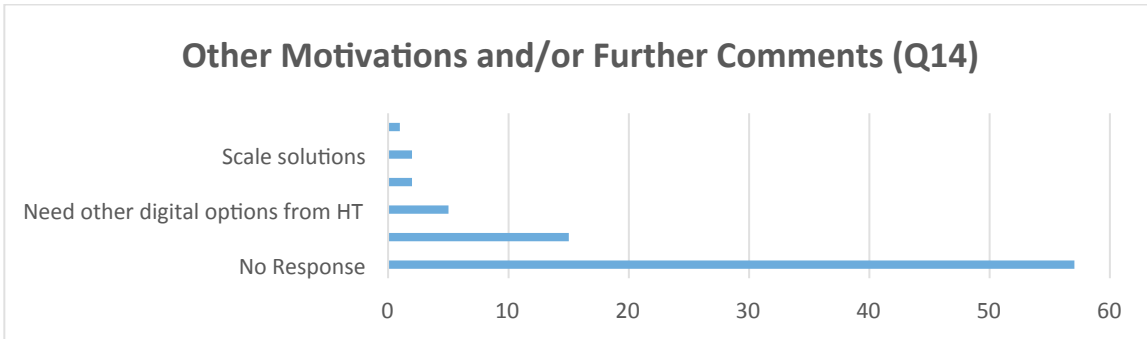


### Question 13

To what degree is your interest in Hathitrust as a collective solution for these newer formats prompted by the following concerns? Scale 1=Not very important; 5=Very important (Q13)							
	1	2	3	4	5	Total	Avg. Resp.
Local infrastructure inadequate	10	15	21	15	9	70	2.97
Local infrastructure too costly to develop and maintain	10	7	23	15	14	69	3.23
Materials are at risk for preservation	5	7	16	25	17	70	3.6
Greater benefit to end users of aggregated access	5	4	7	20	33	69	4.04



**Question 14**



**Question 15**

**How interested is your institution in seeing HathiTrust devote resources and/or development effort to support the following? Scale 1= Not very interested; 5=Very Interested (Q15)**

	1	2	3	4	5	Total	Avg. Resp.
Open Access content development	4	11	20	22	16	73	3.48
Ingesting library published content	10	12	21	24	6	73	3.05
Collaborative Collection Building	4	5	24	20	20	73	3.64
Collection Development Tools and Services	4	8	15	20	26	73	3.77
Eliminating or reducing duplicates within the corpus	16	13	13	22	9	73	2.93
Ability to store high-resolution TIFFs	8	8	14	25	17	72	3.49



How interested is your institution in seeing HathiTrust devote resources and/or development effort to support the following? (Q15)  
 1=Not Very Interested; 5=Very Interested



## Appendix B: Survey Respondents

American University of Beirut	University of Alberta
Arizona State University	University of Arizona
Baylor University	University of California, Berkeley
Boston University	University of California, Irvine
Brown University	University of California, Los Angeles
Carnegie Mellon University	University of California, Riverside
Case Western University	University of California, San Diego
Columbia University	University of California, San Francisco
Cornell University	University of California, Santa Barbara
Duke University	University of Delaware
Emory University	University of Houston
Florida Atlantic University	University of Illinois at Chicago
Georgetown University	University of Illinois at Urbana-Champaign
Harvard University	University of Iowa
Indiana University	University of Kansas
Iowa State University	University of Maryland, College Park
Johns Hopkins University	University of Massachusetts Amherst
Kansas State University	University of Miami
Lafayette College	University of Michigan
Library of Congress	University of Minnesota
Massachusetts Institute of Technology	University of Nevada, Las Vegas
McGill University	University of North Florida
Michigan State University	University of Notre Dame
Montana State University	University of Pennsylvania
New York University	University of Rochester
North Carolina State University	University of South Florida
Northeastern University	University of Texas at Austin
Northwestern University	University of Texas at San Antonio
Ohio State University	University of Utah
Pennsylvania State University	University of Virginia
Princeton University	University of Washington
Purdue University	Utah State University
Smith College	Virginia Tech
Temple University	Wake Forest University
Texas A&M University	Washington University in St. Louis
Tufts University	Yale University
Universidad Complutense de Madrid	

## **Appendix C: HathiTrust Collections Priorities Survey**



## *HathiTrust Collection Priorities*

### **Default Question Block**

As part of HathiTrust's medium- and long-range planning process, HathiTrust would like to better understand member needs with respect to future development of HathiTrust collections and collection-related services. This survey is designed to elicit member priorities for future work to aid in planning future HathiTrust collection development activities. The survey has been prepared by the HathiTrust Collections Committee on behalf of the Program Steering Committee (PSC).

In accordance with discussions at the Fall 2014 Membership Meeting, we consider it a foundational assumption that HathiTrust will continue to build out the current corpus of digitized book content, but anticipate that members may also be interested in exploring support for additional content types as well as other ways of developing value for our collections. Your responses will help to shape future HathiTrust directions in pursuing these collection interests.

The results of this survey will be shared with the Program Steering Committee (PSC), the Board of Governors (BoG), and the HathiTrust membership as a whole, and will provide critical input to discussions about future directions for HathiTrust collections. We appreciate your input into these important discussions.

### **Instructions:**

Please submit one response per member institution. Feel free to collect responses from multiple stakeholders in your institution and combine these into a single survey response. Please submit your response by October 26, 2015. Questions about the survey can be sent to: Claire Stewart, University of Minnesota, [cstewart@umn.edu](mailto:cstewart@umn.edu) You can download a blank copy of the survey here: <https://umich.box.com/s/rbk1ioog3yxusp3xh8s6gfkf0oi7mt4d>

**Institution Name:**

**Respondent Name:**

**Respondent e-mail**

1. Has your institution contributed content to HathiTrust to date?

- Yes
- No

2. If Yes: Approximately how many volumes have you contributed as of June 2015?

- 1-10,000
- 10,001 - 100,000
- more than 100,000

3. Do you anticipate contributing volumes to HathiTrust in the next year?

- Yes
- No

4. Approximately how many volumes do you expect to contribute in fiscal year 2015-16?

- 1-10,000
- 10,001 - 100,000
- more than 100,000

5. If you have not contributed content to HathiTrust, please tell us why. (Check all that apply)

- We have not digitized any books
- Our digitized collections are out of scope for Hathi
- We prefer other solutions
- We didn't know that we could contribute content
- We tried, but could not meet HathiTrust specifications
- Staffing constraints have made this infeasible
- Other (Please Explain)

6. Are there locally digitized collections that you have not been able to ingest in HathiTrust but would like to? Please describe.

### Notable Collections

For the purposes of this survey, we are defining Notable Collections to mean collections in a particular subject or domain that you consider to be nationally significant.

7. Please describe any notable content your institution or organization has deposited in HathiTrust to date:

8. Please describe any notable content your institution or organization is planning or would like to deposit into HT in the next 1-3 years (both scope/volume of materials and

descriptive attributes)

### Collection Prioritization - Current Collections

9. How important is it to your institution that HathiTrust continue to develop and expand the current print-based book corpus?

	1	2	3	4	5	
Not at all Important	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very Important

10. How would your institution prioritize the following strategies for further developing the book corpus:

a. Quality improvement of the existing corpus

	Not at all Important				Very Important	
	1	2	3	4	5	
Missing foldouts				<input type="radio"/>	<input type="radio"/>	
Missing, illegible, or out-of-order pages				<input type="radio"/>	<input type="radio"/>	
Gaps in multi-volume serials / sets				<input type="radio"/>	<input type="radio"/>	

b. Addition of new material

	Not at all Important					Very Important
	1	2	3	4	5	
Mass digitization rejects (pamphlets, oversize, brittle materials, etc.)					<input type="radio"/>	

- Digitization of in-copyright printed books
- Open access books
- Born digital monographs (regardless of rights status)
- Targeting specific subjects
- Targeting specific corpora recommended by scholars or other expert groups
- No new strategies needed, continue to build via independent effort as now

c. Other (Please describe)

### New Collection Priorities

Any expansion of HathiTrust collecting priorities into new areas is likely to involve an additional cost. With this in mind, please describe your institution's interest in HathiTrust as a resource for the following types of materials.

11. How interested is your institution in using HathiTrust for *preservation* of the following content types?

	Not at all Interested			Very Interested	
	1	2	3	4	5
Audio materials				<input type="radio"/>	<input type="radio"/>
Moving images				<input type="radio"/>	<input type="radio"/>
Still Images				<input type="radio"/>	<input type="radio"/>
Ephemera				<input type="radio"/>	<input type="radio"/>
Manuscript and archival materials				<input type="radio"/>	<input type="radio"/>
Maps (sheets and sets)				<input type="radio"/>	<input type="radio"/>
Encoded texts				<input type="radio"/>	<input type="radio"/>



Executable content	<input type="radio"/>	<input type="radio"/>
Web archives	<input type="radio"/>	<input type="radio"/>

Other (Please Describe)

12. How interested is your institution in using HathiTrust to *provide access* to the following content types?

	Not at all Interested			Very Interested	
	1	2	3	4	5
Audio materials				<input type="radio"/>	<input type="radio"/>
Moving images				<input type="radio"/>	<input type="radio"/>
Still Images				<input type="radio"/>	<input type="radio"/>
Ephemera				<input type="radio"/>	<input type="radio"/>
Manuscript and archival materials				<input type="radio"/>	<input type="radio"/>
Maps (sheets and sets)				<input type="radio"/>	<input type="radio"/>
Encoded texts				<input type="radio"/>	<input type="radio"/>
Executable content				<input type="radio"/>	<input type="radio"/>
Web archives				<input type="radio"/>	<input type="radio"/>

Other (Please Describe):

13. To what degree is your interest in HathiTrust as a collective solution for these newer formats prompted by the following concerns?

	Not at all Important			Very Important	
	1	2	3	4	5
Local infrastructure is inadequate				<input type="radio"/>	<input type="radio"/>
Local infrastructure					

- would be too costly to develop and maintain
- Materials are at preservation risk
- Greater benefit to end users of aggregated access

14. Other motivations and/or further comments:

15. How interested is your institution in seeing HathiTrust devote resources and/or development effort to support the following:

	Not Very Interested			Very Interested	
	1	2	3	4	5
Prospective Open Access Content Development (e.g. hosting new OA material such as Knowledge Unlatched, cultivating other new OA partnerships)				<input type="radio"/>	<input type="radio"/>
Ingesting Library Published Content				<input type="radio"/>	<input type="radio"/>
Collaborative Collection Building (e.g. building collections in particular subjects or formats through intentional collaborative effort)				<input type="radio"/>	<input type="radio"/>
Collection Development Tools and Services (e.g. collection analysis, defining and presenting collections within the HathiTrust corpus)				<input type="radio"/>	<input type="radio"/>
Eliminating or reducing duplicates within the corpus				<input type="radio"/>	<input type="radio"/>
Ability to store high-resolution TIFFs or similar for preservation purposes				<input type="radio"/>	<input type="radio"/>

Other (describe)

16. Please feel free to describe or elaborate on your interest here:

### 17. Additional Input

HathiTrust values your additional feedback about improving our collections. Please use the space below for any further information or comments you would like to share.

Please email [Melissa Stewart](#) if you have any questions or problems with this survey.



## Appendix D: Open-Ended Responses

Of the 76 institutional respondents to the survey, 69 provided text responses to some/many/all questions that offered open-ended “Other” or “Describe” prompts. These responses have been slightly edited to anonymize institutional identity. While these responses have not been coded for further analysis, they are included here as an important part of the survey response data set.

---

### HathiTrust Contributor Status

**Question 5:** If you have not contributed content to HathiTrust, please tell us why. (Check all that apply)

Other:

- We are planning to digitize 5000 out of copyright, pre 1956 Arabic books.
- It has been only recently that [we] digitized collections appropriate for HathiTrust; other digitized collections are in our institutional repository.
- Lack of interest
- Changes in personnel assignments have now made it possible to consider and plan for adding content to the HathiTrust.
- don't own notable content
- Still developing digital preservation priorities, focusing on our own institutional repository
- We have not looked into the process or HT specs
- Our digitized books are mostly in Internet Archive
- We're new members, are not doing much digitizing yet, but that will likely change over the next few years.
- No workflow established to meet HT deposit requirements
- We're working on contributing content we have deposited in the Internet Archive
- This is our first year as a HathiTrust member
- Constraints of original agreement with Google Books partnership
- Our digitization efforts have mostly been limited to unique items that may be out of scope for HathiTrust.
- We have digitized books as part of the Religion in [our] LSTA grant but have not ingested them yet.
- We have to do custom coding and processing of MARC records to meet the HathiTrust specifications

---

**Question 6:** Are there locally digitized collections that you have not been able to ingest in HathiTrust but would like to? Please describe.

- Yes, the Israeli Pulp fiction collection -- see comments under Notable Collections.
- We would like to submit our nineteenth century [literary collection], but we are only able to submit access-quality images, not hi-res preservation files for 75% of the collection.
- We are trying to submit collections locally digitized and/or digitized for us by vendors which we have uploaded to Internet Archive. We are trying to submit them to Hathi from IA and are running into difficulties, apparently for a variety of reasons which are not easy to diagnose. It would be very helpful to have straightforward written instructions on file specifications, metadata requirements, and procedures.
- Yes. Recent experience with collections in our legacy [digital collections] system has led us to migrate content to new in-house local HYDRA repository, due to various factors (copyright status that would adversely affect full-view ability in HT, content type (archival; mixed content collections) and lack of local MARC)
- We have some uncataloged materials in Internet Archive (e.g., city directories) that we would like to ingest into HathiTrust, but we do not have the cataloging resources available to create the required records.
- 1. "Bound withs" and "Bound together" have been a problem, since they require each submitted "volume" to have a unique identifier. Note: given our local practice, it is difficult to assign barcodes to each bound-with part. 2. Uncataloged bound volumes (ofc there would have to be some metadata, but Hathi requires not only MARC records but also OCLC numbers) 3. Non-book formatted materials. We regularly digitize non book formats, such as AV materials and photographs. Rights issues have prevented us from actively considering contributing such content to a digital library. It's possible that this could be a future consideration once we have developed and implemented a local strategy for rights and permissions.
- We do not know this at this time. This is something that we will need to assess.
- Most of our [literary and] e-text projects.
- We have notable collections that have been digitized, in particular [institutional] serial publications. Some of this content has been added by other institutions but is spotty. We could provide complete runs and give OA permission for post-1923 publications. We also have Iowa history and other monographs but have not checked to see if they already exist in HT
- Digitized books - but it is not your fault, it is ours.
- Books about [our university and state publications]
- We have not moved the balance of the ~30,000 volumes that were digitized and are in IA while we sort out some metadata issues in our ILS records. Also, the only way that we know of for LC to load content into HT is via IA and at present we are not using IA as a scanning contractor. We have a variety of book-type bound materials that could be in HT that have been digitized at LC. In general we are reassessing our public domain book and serial digitization needs.
- Yes. We are currently focused on a [institutional] theses digitization project and some individual rare book projects. We have analysed our collections with Sustainable Collections Services and should have 100,000 titles in the public domain that aren't yet in HathiTrust, so longer term we would like to add all of those.

- We have a spoken word collection, and some of that content has been digitized. We would be glad to see a place for that material in Hathi.
- Does this refer to straight text or more varied collections?
- Perhaps in the future...
- We would like to send a copy of all books digitized prior to current workflow (which includes sending a copy to Hathi). This is about 4,000 titles, though some may be duplicated in Hathi. Also, there are a number of foldout pages from environmental impact statements we've digitized locally and would like to send, pending some local metadata work that needs to be completed.
- Historical [state]-related materials, brittle books, etc.
- Yes. We have digitized roughly 1,000 volumes (mostly public domain) in house over the past 10 years as part of our brittle books program. We have made PDFs available through the Internet Archive, and have page image tiffs stored locally. Some of these titles may have been digitized elsewhere and deposited to HathiTrust in the interim, but we have not yet prioritized finding out which ones nor begun preparing them for deposit.
- Yes, journal titles and issues in our [specific] Project. We also plan to upload journals from our [specific] Periodicals Project up to the 1970 revolution and from our collaborative project with [another institution].
- We have been digitizing many [institutional] agricultural extension publications. There would probably be interest in adding these to HathiTrust, but we'd need to have some conversations about that locally first.
- Too early to tell.
- Not all local collections have page images that could be sent, so those items can't go to HT.
- We would like to contribute a portion of the [state] Sanborn Fire Insurance maps that are in the public domain.
- We route all our locally digitized books through Google for inclusion in Hathi. We would very much like a streamlined process for ingest into Hathi for our materials that we digitize that Google can't.
- Yes, we have many. We need an clear workflow beyond replacements that we have worked out with both CDL and HT.
- Other local digitized image collections
- Historic Cartographic material , 19th Century Prints, 14th-18th Century Manuscripts , Scientific Film recordings, Historic Photographs, Academia Drawings, Engravings
- Possibly. We have some in-house digitized material, approx. 10,000 book-like objects and 100 historical newspaper titles totaling 700K pages.
- We would most likely identify the unique book/published materials holdings we have and consider submitting those materials to HT.
- We have attempted to use the Brittle Books feature meant to provide institutional access to in-copyright works digitized due to their brittle condition (<https://www.hathitrust.org/out-of-print-brittle>), but have yet to see it work. In the best of circumstances there is significant lag between deposit and access. We are disappointed to have to consider a local solution for these volumes.

- Collections that contain structured text, including journal collections with article-level markup, as well as fully encoded text collections.
- We have a backlog of more fragile items that were rejected by Google but we are working with a vendor to digitize them anyway. This will include up to 1,500 volumes of materials not currently in Hathitrust and not widely held by other institutions.
- No, the ability to ingest hasn't been the problem for us.
- We just became a member of the HathiTrust through [a] consortium.
- If we could be reminded of HT's collecting scope, we can review our digitization plans and determine whether we may have items that would qualify.
- I would love to see us finish digitizing our pioneer diaries and get them ingested. I also think it would be great to have HT archive our historical newspapers.
- Yes, and as new depositors to Hathi we have a backlog of files to submit.
- We only have digitized unique content.. Photographs, oral histories, manuscripts
- [State] Agricultural Extension publications, public domain culinary history collection
- Yes, and we will, once we have an established local workflow.

### **Notable Collections**

**Question 7:** Please describe any notable content your institution or organization has deposited in HathiTrust to date (Note: an effort to anonymize is not provided for the following responses, which describe collections publicly known to be available in HathiTrust):

- Our contributions to the Medical Heritage Library (<http://www.medicalheritage.org/>)
- All we have deposited into HT are Google and MS/Kirtas scans to date. Local workflows to HT are forthcoming.
- Confederate Imprints; Jantz Collection (German Baroque Literature; German Americana); Utopian Literature; Ottoman Turkish Collection; Emblem Books
- 1. Medical Heritage Books (about 30 volumes) 2. African American Imprints (the bulk of our contribution so far) 3. Southern Imprints
- This is content that was part of our Google Books scanning project, and we feel it is all nationally significant.
- Folklore Collection
- LC to date has only loaded into HT books that were digitized by IA from the LC general collections that by definition are not "notable." We have strong Canadiana content, because we are one of the oldest universities in Canada and have strong collections related to the the history of Quebec and Canada. We also have various special collections related to other subjects, like Voltaire, the Age of the Enlightenment, the Burney family (Charles and Fanny), the history of medicine and William Osler, children's literature, and others.
- Some of our strengths are in Africana (African Studies), and agriculture (including areas such as Turfgrass and veterinary medicine). Some of this material appeared on our picklists for scanning by Google.



- Botany and entomology collection
- Transportation (approx. 79,000 items) Medical (approx. 25,000 items) Music (approx. 10,000 items)
- We have deposited more than 100,000 Google-digitized volumes as of October 2015, mostly in the public domain, including 4,965 Charvat (popular American fiction) items; 1,508 items from the Billy Ireland Cartoon Library and Museum – much of this is periodicals; 621 Theatre Research Institute items.
- All of our Google Books Project content.
- none. Almost all of our previous deposits were part of the CIC/Google government documents project.
- A collection of the entire legacy run of the Bulletin of the Texas Agricultural Experiment Station (1888 to 1998)
- 1,463 volumes of the John A. Seaverns Equine Collection, part of the Webster Family Veterinary Library Special Collections, Institution #62. All volumes were digitized by the Internet Archive. Subject areas are related to all aspects of horsemanship, representing the lifelong collection of John A. Seaverns. The collection is especially strong in racing, hunting and equestrian art.
- More than 3,000,000 items from high density storage which includes some notable materials. Currently digitizing 40,000 dissertations of Institution #37 for Hathi (via Google). Strategic digitization of federal government documents. Large collection (40,000) sheet music collection.
- Outside of the established CDL workflows, Institution #38 has set up a replacements workflow but otherwise we do not have the workflow established for local (read non Google or IA scanning) to HT. We are eager to establish workflows that would provide an easy path to local digitized notable content and collections into HT.
- Scripps Institution of Oceanography largely scanned by Google, but we are adding local scans of the rare books from this collection as a current project.
- Incunabula Complutense. This collection contains a selection of 15th century printed books held at the University Manuscripts . 78 manuscripts (11th - 16th ) held at the University
- Corks & Curls, a number of Commonwealth of Virginia and [institutional] historical documents
- The Canadian Institute for Historical Microreproductions (CIHM) monograph collection.
- UCSF University Publications: this collection contains materials published by university schools, programs, and research institutes (course catalogs, announcements, student publications, annual reports, newsletters, etc.) as well as yearbooks dating back from 1864 held at the institutional Archives.
- Contributed to the development of the shared collections of the University of California, from which much notable content was deposited in HathiTrust.
- Nothing of any great note so far.
- We have large digitized collections of images from our Special Collections, primarily photographs. There are other digitized documents and relatively small numbers of audio and video / film items.

- In addition to significant amounts of Google-digitized content now flowing, the following: Emblematica, Illinois History, Illinois state documents, and materials related to NEH microfilming projects.
- Currently contributing via Google Books Project.
- We deposited Hebraica materials.
- Much of what we have contributed is the digitized copies of our print Massachusetts State Documents.
- We have digitized broadly, rather than deeply. Modest portions of notable collections are included, including Transportation History, Philippines, Islamic Manuscripts
- Through the Google Book initiative: US Government documents; materials related to forestry, entomology, Scandinavian area studies, language and literature.
- So far, only brittle books scans.
- None
- We have loaded digitized books from a Microsoft funded digitization project.

**Question 8:** Please describe any notable content your institution or organization is planning or would like to deposit into HT in the next 1-3 years (both scope/volume of materials and descriptive attributes)

- The [specific] Pulp fiction collection is the sole compilation of [specific] Pulp fiction in the United States and addresses several aspects of [specific] culture. To date, 467 volumes are digitized, which is half of the current collection. [The] Libraries continue to add to the collection.
- [Specific Publication] (1,130 issues) The [Specific Publication] began publication on October 28, 1897 in [specific city and state]. The [institution's] collection includes the full run of the [Specific Publication]. Eight-hundred and ninety-nine writers contributed 4,349 articles, and a qualified editorial staff prepared the manuscript for publication.
- No planned deposits at this time; however, this could change.
- We have a small collection of titles on the history of science that are significant. The donor used Dibner's Heralds of Science as his collection list. We have randomly checked the HT collection and think we have some unique content to upload.
- Public domain published materials from the [specific] Library. Public domain pamphlets from the [specific] Library
- [Institution's] part of MOA - significance is an early digital preservation project will be brought into current preservation practice through deposits exemplifies ongoing curation of assets. Development of local workflows that deposit locally digitized items directly into HathiTrust will allow us a way to deposit thematic collections, and digitize rare and unique materials for presentation.
- We expect to continue building on the collections mentioned above and possibly expanding content from our History of Medicine Collections. Scope and volume uncertain at this point.

- 1. Triple Deckers (382 volumes from Internet Archive (IA); 234 vols not yet digitized) 2. Baedekers (145 volumes from IA) A collection of Baedeker European travel guides from the late 19th century. 3. Regimental Histories (379 volumes from IA; 43 vols not yet digitized) 4. American Methodism (1,019 volumes from IA; 1,500+ volumes from Lyrasis digitized microfilm). 5. Atlanta City Directory (25 volumes from IA) 6. African American Imprints (179 volumes from IA; 481 vols not yet digitized) 7. Yellowbacks (1,241 volumes from IA; 557 vols not yet digitized) 8. Medical Heritage (192 volumes from IA from IA) 9. Southern Imprints (?? volumes from IA from IA; 2803 vols not yet digitized from all campus libraries)
- We need to investigate this across [our institution], and do not have the information at this time.
- Late 19th C early 20th C Near Eastern collections
- [Specific] Archives.
- We do not have plans to do so.
- [Our library] holds tech reports, proceedings and other materials that are not widely replicated in other institutional collections. A recent collections analysis suggests that [our library] owns roughly 200,000 items for which there are fewer than 3 holdings in the U.S. Nearly all of these items are not currently in HathiTrust.
- Our focus this year is a Canadiana collection called [specific name] that we hope to add about 3,000 titles to HathiTrust for in the next 1 to 2 years.
- We would like to deposit audio files, possibly moving image files, and potentially materials from our American popular culture collection (such as comic art and comic books, which were not picked up by Google due to size and publication format). Even a small effort as proof of concept would be interesting, and might shed light on how scalable these projects would be.
- 1. [Specific Title] Collections Online is a publicly available digital library of public domain [non-English] language content. This mass digitization project aims to expose up to 15,000 volumes over a period of five years. 2. The [Named Country] Digital Library was created to retrieve and restore the first sixty years of [the named country's] published cultural heritage. The project collected, cataloged, digitized, and has made available over the Internet as many publications from the period 1871–1930 as it is possible to identify and locate.
- Environmental impact statement foldouts (mentioned above) University Press open access books: Would like to make open-access books from our university press available through Hathi, but the format for files (as required by Mellon grant supporting this project) must be epub3, a format Hathi does not support.
- Rare books relating to Irish literature, Italian literature, Medieval studies, Theology and the history of Catholicism in the US.
- •Up to 2,500 rare items from area studies collections such as Ottoman Turkish, early Arabic, Persian and Chinese books, and historic Jewish pamphlets. Many of these titles were requested for scanning through the Google Books project but had condition and/or format restrictions (e.g. brittle paper, tight bindings) that made them poor candidates for mass digitization. They have high research value nationally and internationally, represent

a mix of monographic and serial content, were published in a range of non-Roman languages, and are mostly free from copyright restrictions. •Up to 1,000 additional rare and unique items from [named] special collections. •Up to 150,000 state and federal government documents through Google Books. •Up to 15,000 volumes published between 1923-63 that are out of copyright through Google Books.

- The [institution's] agricultural extension publications mentioned in Q6 would be candidates. There are potentially other textual materials as well that we plan to scan over the next few years,
- For next 1-3 years: Collection of the entire legacy run of the Bulletin of the [specific state's] Agricultural Extension Service (1914 to 1998) (about 700 items) Collection of the entire legacy run of the Leaflets of the [specific state's] Agricultural Extension Service (1914 to 1998) about 2459 items)
- Notable content that we would like to contribute are a portion of our locally digitized maps collection, namely the [specific state's] Sanborn Fire Insurance maps dating from the late 1880s to 1923. The Sanborn Map Company, the best known of the US fire-insurance map producers, has made maps since 1867. The fire insurance maps produced by Sanborn show building footprints, building material, height or number of stories, building use, lot lines, road widths and water facilities. The maps also show street names and property boundaries of the time. This collection of maps is historically significant as it is sometimes the best detailed map of a town or city dating from the mid 1800s.
- (a) 25,000 pre-1923 items in special collections; (b) a large historic folio collection across all the subjects and campus libraries; (c) a unique 18,000 libretto collection
- Some examples: Rare Hebraica volumes held no where else in the world, Rare Turkish Satirical Periodicals, Rare Urdu content and collections
- [Italian artist's] engravings. 182 [Italian artist's] engravings. The complete collection of [Italian artist] is composed by 1135 engravings of him and his son. Engravings of [specific 1st century scientist's] collection: Collection of 50,000 prints and illustrations contained in the 3,000 digitized books in [specific 1st century scientist] project. [Several other special collections across the ages, including image collections, are detailed.]
- The Google Book Project digital files which I hope can be added to HathiTrust this year.
- Western Canadian material. 10K monographs, 650K newspaper pages, etc.
- Through the Google Books project, we will contribute approximately 16,000 items including [Specific] State Documents, Special Collections Materials, and out-of-copyright materials from our general collections.
- [Specific Medical] Collection of pamphlets: The [specific medical] collection contains over 600 works related to cholera, epidemics, and public health and includes books, pamphlets, government reports, letters, and manuscripts, from the 18th to the 20th century. Among the rarest items in the collection are 265 pamphlets and reports that will be submitted to HathiTrust. Collection of State Medical Journals (1900-2000): The [specific] Library is collaborating with four other medical libraries on a project to digitize and make publicly accessible state medical journals from 1900 to 2000. As a result of this project [specific] library will digitize and upload about 850 volumes to HathiTrust.

- [Specific] State House & Senate Journals
- Our Rare Book and Manuscript Library is depositing volumes. They are happy to continue to deposit books here, but frequently say they would prefer an interface for rare materials that better emphasizes their special qualities.
- There are no current discussions underway to contribute additional content beyond Google Books Project which includes monographs, US government publications and Iowa state documents.
- We are looking at Persian and Chinese language materials. We do not know the number of volumes at this time. Some of it is already digitized, some of it is in the process of being digitized.
- [The specific institution's] Press backfile (approximately 300 volumes) - output of the [specific institution's] Press where we hold copyright - Serials and monographs from our [specific] Heritage Collection (estimates forthcoming)
- Possibly the [named] Sheet Music Collection, beginning with approx. 40,000 pieces published before 1870 -- we are awaiting word from a potential funder for this project.
- We would like to add many of the published textual materials now held within our [image and real-time media repository].
- We have digital images for local special collections related to (1) horses, (2) World War II pamphlets, (3) Japanese juvenile novels, and (4) medieval manuscripts that we plan to contribute eventually.
- Latin American books and serials (approx. 500,000)
- Five thousand rare book scans of materials in the National Central Library of [East Asian country]; as well as smaller numbers of other international studies materials.
- [Specific state] Elusive Documents collection: The Elusive Documents collection consists primarily of publicly funded scientific reports that, for whatever reason, received a limited distribution upon publication. These reports are of significance either to the State, region, or to the university curriculum. The primary area of emphasis is on the broad subject of natural resources. Scope: several hundred documents and growing
- Religion in [specific state]

## Collection Prioritization – Current Collections

**Question 10:** How would your institution prioritize the following strategies for further developing the book corpus: Sale of 1 to 5, with one being not at all important and 5 being very important (Numbers Listed here are Averages)

Other:

- We were not sure how to rank the last item. We think continuing current strategies is very important, but also feel the new areas listed above have importance as we coded them.
- More digitized out of copyright materials

- One of the critical functions that Google has played in building the HathiTrust corpus is deduplication: partner libraries submit their holdings and receive picklists of items that Google has not yet digitized. Without Google to compare members' holdings and request as-yet-undigitized items, responsibility for making the determination of whether a given item has already been digitized (and thus whether to expend local resources to digitize it and contribute it to HathiTrust) transfers back to the libraries. This has been a challenge from the start for any HathiTrust members that are not also Google Books scanning partners. But, it will become magnified as Google Books scanning winds down and member libraries ramp up their in-house and vended book digitization efforts. At this point, decisions about which volumes to scan are primarily being made locally on an item-by-item basis, but this is very time consuming and will not scale.
- More U.S. government publications (not just current gaps in multi-volume sets, but also titles that haven't been digitized yet).
- Our highest priority is for stronger quality control and assurance. Completeness is an important component for reliance upon the digital copy. Gaps in multi-volume serials/sets prevents complete discovery and context of the collection as a whole. Accessibility for the print disabled is of great importance so the digitization of in-copyright printed books is still a high priority. There's high value to discovery and access to commonly-held collections; discovery to readership.
- Give serious consideration to scanning outputs from disability services. FYI, the "born digital" item seems problematic in its characterization.
- Endangered-language materials; nonmember collections
- We would rate adding additional public domain (pre-1923) books that are not already in HathiTrust as a very high priority.
- Data mining
- Federal documents

---

## New Collection Priorities

**Question 11:** How interested is your institution in using HathiTrust for *preservation* of the following content types? Scale of 1-5, with 1 being not very interested and 5 very interested (Numbers listed here are Averages)

Other:

- Newspapers at level 3
- Emails, Tweets
- We are having discussions around preservation needs and have not yet fully explored various options, assessed costs, or explored solutions. Therefore, at this time, we're neutral regarding our interests in using HathiTrust as a preservation solution.

- Types ranked as 4 would be higher if MARC metadata were not a requirement for ingest, or assuming that collection-level records would make for a satisfying user experience. AV ingest would depend upon costs relative to other solutions.
- 

**Question 12:** How interested is your institution in using HathiTrust to *provide access* to the following content types? Scale of 1-5, with 1 being not very interested and 5 very interested (Numbers listed here are Averages)

Other:

- Newspapers at level 3
  - HT, let's have snippet access to all HT content!
  - Ephemera is of particular importance for research of and for underrepresented populations.
  - Access to AV might require more finely-grained authorization. Material types rated 4 would be higher if the MARC record and presentation issues noted in Q. 11 were resolved. Costs would need to be compared to other solutions.
  - The Internet Archive seems to be covering nonprint fairly well.
- 

**Question 14:** Other motivations and/or further comments:

- We see the advantages of working to scale even if our local infrastructure is adequate
- We participate in the CDL's Web Archiving Program; we are currently transitioning to the Internet Archive's Archive-It Service from CDL.
- One of our motivations is that HathiTrust is a COLLECTIVE solution with access.
- Support for new needs such as accessibility and text mining: see additional comments in Q17.
- Involved in other project addressing audio and video content
- Answers vary by content type
- We prefer that HT keep its focus on print resources.
- Hathi is still figuring out how to do books and serials well and should focus on that. Maps are a common foldout in books and it would make sense for Hathi to solve the larger problem of maps at the same time they solve the foldout problem.
- A preservation solution is still yet to be determined.
- I hope it's obvious that the real interest is in providing much higher quality services and access at much lower cost. Even if we were more generously funded to do something local, doing so would not advance our library's interests as much as leveraging HT and doing more appropriate local things.
- Interested in solutions at scale
- Interesting solutions for AV and web archives are developing outside HathiTrust.

- Regarding manuscript and archival material -- DPLA seems better positioned now for these formats.
  - It is our opinion that printed (paper-based) materials are the best fit for HathiTrust to provide access to, although bit-level preservation for digital materials is still desirable even without access.
  - At this point, we are not interested in pursuing these additional formats.
- 

**Question 15:** How interested is your institution in seeing HathiTrust devote resources and/or development effort to support the following: Scale of 1-5, with 1 being not very interested and 5 being very interested (Numbers listed here are Averages)

Other:

- IIIF viewer compliance
  - Improved discovery, especially of serials, and integration with library discovery platforms
  - Standardize ordering of serials holdings within the HT discovery interface
  - Non print formats
  - We are most interested in having HathiTrust continue with its original priorities.
- 

**Question 16:** Please feel free to describe or elaborate on your interest here:

- [We] would like HathiTrust to store and preserve high-quality enhanced files, and to provide access to those files. We also encourage HathiTrust to digitize and make accessible in-copyright print books, and thereby help prevent the information in these books from disappearing as we move further into the digitized future.
- We have interest in the ability to store high-resolutions TIFFs for preservation; however, the limited number of our collections that meet the HathiTrust collection development policy diminishes the value of this for us.
- We are in the stage of considering digital preservation options and we might consider what HathiTrust has to offer in this realm. We have digitized books in the past - we've put a little over 1800 books into the Internet Archive.
- We're generally interested in collaborative collection development, but we have questions about how that would work in a HathiTrust context.
- We are only interested in the ability to store high-res TIFFs for preservation if the price is cost-effective as compared with other solutions.
- We aren't sure what you mean by "Ingesting Library Published Content," but we are guessing that you mean ETDs.
- Very interested in digitized ephemera.
- It's getting increasingly important that in cases of titles with multiple years/issues/parts that the list of items, especially for full text, appear with similar tags/labels and in



chronological and/or numerical order. We've embarked on a project to reduce our print and microform government documents collection based in part on availability in HathiTrust, and it's very difficult to determine if a run is complete or not when there are dozens of entries in apparently random order (probably in the order ingested, but that can be pretty random sometimes).

- The potential around collection analysis tools is intriguing if we believe that Hathi could provide something more useful and as a competitive cost as compared with other solutions. Without that it seems like a poor use of resources. On the whole perspective of open access, incorporation of born digital where it fits into the broad mission of preserving important materials seems a good fit with Hathi's current mission. Hathi could take on the role of a LOCKSS and Portico-style platform and preserve access to OA resources that may not be on the radar of other platforms.
- We are very interested in data and text mining capability development. We would also like to see efforts dedicated to quality enhancements to the corpus :)
- In reference to responses in #15: Collaborative Collection Building - lower interest due to the idea that HathiTrust is a repository of the broad cultural record rather than become a site to go to for particular subject emphases. Collection Development Tools and Services - lower interest due to not knowing the purpose of the analysis; what impact would it have; how would it affect decision-making? Eliminating dups - do duplicates affect text-mining and search results? If so, this is of high interest. What are the problems caused by duplicates?
- When searching a serial title in the HT, the holdings do not appear in any sensible order from a content perspective (perhaps in load order?). It would make much more sense for us at the library, as well as users, to have holdings of a multivolume set appear in date/volume order.
- [We] Would like to see HathiTrust support collaborative/cooperative print retention initiatives across member institutions.
- Standardized ordering of serials holdings is necessary to provide users with an easier way to interpret what is held in HathiTrust.
- I'm not sure we have a huge amount of interest in seeing HathiTrust develop a suite of tools. (We don't object in principle, but we recognize that such development would be underwritten by membership fees and it's not necessarily something we're interested in paying for.) HT's value to us -- which is tremendous -- is mainly as a repository of fully-searchable and downloadable texts. Again, I don't want to give the impression that we oppose this kind of development, but there's a limit to what we want (and can afford) to pay for.
- We are looking for ways to rely on Hathi as surrogates for physical volumes in our collection, especially for monographs, as an aid to reducing our physical collection footprint.

**Question 17:** HathiTrust values your additional feedback toward improving our collections. Please use the space below for any further information or comments you would like to share.

- We're more interested in providing better access to the existing HathiTrust scope rather than expanding it.
- Interesting tension between research with the collection and teaching with the collection. How do our communities make sense of the HathiTrust collection and can that inform collection choices? Are there uses that we have not yet imagined? One hopes that the HTRC can help with some of that. A huge challenge ahead is the task of prospectively developing the collection as titles , as works are released. What does it mean to ingest born digital items when those items may be multiformat or have digital objects embedded in them?
- We are concerned that HathiTrust be very deliberative about expanding to such an extent that the price becomes prohibitive for some libraries, especially since various of these proposed services may not be of high interest to some institution. There may be a need to use a somewhat different pricing structure for some of these, so that they are "optional" rather than spreading out the cost completely among all institutions. We would not want to find ourselves priced out of the current books corpus because of the requirement to pay for other services that are not of compelling need for us.
- Some quality control concerns about the content and the metadata.
- [We are] a new organization and in the process of understanding its partnerships; we are in the process of thinking through the various strategic partnerships we have and want to build. Our work over the next few months in this arena will help us to more clearly define all those partnerships. We intend HathiTrust to be a partner.
- Suggestions - actively pursue permission to open up post-1923 titles, improve metadata quality, and clean up holding statement chronologies.
- Extremely interested in digitization/preservation of GPO materials and other government publications.
- We would be interested to see how Hathi could support some emerging needs. One is support for persons with disabilities (vision, hearing, etc.): providing OCR text for example. OCR text is also relevant for a second interest: continuing to grow the role of Hathi to support text mining projects for faculty and researchers. Third, we would like to see expansion to cover audio and moving image materials: this could include (and perhaps might have to include) transcripts and similar supporting elements, which are needed for ADA compliance but would also assist a wide range of users.
- Really wish HathiTrust made it easier to upload local content into the corpus and had more support staff to help.
- [Our] Archivist comments "If HathiTrust decides to begin including archival/manuscript material that would be very helpful in bringing together unique content of high research demand into one online location." Thank you for offering this opportunity for input.
- Highest priority is to improve the completeness and level of quality of the current collections. Make this text corpus more usable for advanced research, such as text mining. Possibly expand into OA textual-type materials.

- [We] would like to see HT ease the burden re establishing and promoting ADA and print disabled access more broadly. The current process is too burdensome and is not making it easy to promote the services. Many thanks for the opportunity to comment.
- Encourage further work to be done around international copyright support and brightening public domain materials, e.g. Canadian federal government publications.
- Services for persons with print disabilities remain important.
- We continue to be disappointed by the hurdles presented to users with print disabilities, esp. when the courts have made clear the role and opportunity. Requiring a mediating agent for these individuals is separate and not equal treatment. A more effective strategy could also leverage broad interest in digitization in support of these individuals, thus also improving the HT corpus and value to our community.
- For rare books: images within books and including information about photos, tables, frontispieces, maps, etc. Spines of rare books. Ability to search by size from the database. Frontispieces & maps. Can HathiTrust metadata be organized such that it plays as well with discovery systems, such as Primo Central as commercial vendors do? It would be great if HathiTrust materials came up in search results from the Libraries' homepage.
- As we collectively begin to look closer at our collections, to compare them with the holdings of others and to digital surrogates, we will discover far more unique copies than previously suspected, including among those currently identified as duplicates. HathiTrust can provide assistance with disambiguation by providing checklists for comparison of apparent corpus dups or for comparison of corpus copies with print copies held by participating institutions; tools for updating (or suggesting moderated updates to) existing metadata, including addition of copy-specific details about specific scans in the corpus.
- HT remains one of the best things that has happened to our library in the last 10 years. Thank you.
- HathiTrust has a critical role to play as academic research libraries work to develop collective collections and tackle shared problems in a variety of areas including shared print archiving and text & data mining. I hope to see my library increase its engagement with HathiTrust over time to collaborate on shared challenges and provide improved services to our users.