



HATHI TRUST

What You Want to Know Answers from HathiTrust Staff

This report includes HathiTrust staff answers to questions submitted during the registration process for the 2019 Member Meeting. HathiTrust asked future attendees, “Based on the theme, what question would you like answered at the meeting about HathiTrust's work towards “Driving and Supporting Change”?”

These questions covered everything from international membership and diversification of the collection, to ROI and IIF, repository infrastructure and HathiTrust relevance to open scholarship. While we addressed many attendees’ questions during the meeting, the range of interests was too broad to allow adequate attention to answers during the meeting. Please contact us with any questions about the report. (feedback@issues.hathitrust.org)

[Membership](#)

[Open Infrastructure and Scholarship](#)

[Collections: Diversity and Inclusion](#)

[Technical Infrastructure and Interoperability](#)

[Access and Formats](#)

Membership

- **What developments are underway (governmental, outreach, other) to develop an increasingly international partnership?**

In the last several years we have seen a surge of interest in membership from non-US institutions. In 2018 and 2019 we added 28 new members, 3 of which are located in Canada and 3 in Australia. In 2019 we adopted new, formal membership criteria that clarified that membership is open to non-US institutions.

December 5, 2019

Prepared by HathiTrust Staff



When evaluating how we might continue to expand the membership, HathiTrust and interested libraries must weigh multiple factors when expanding international membership. Top considerations are whether HathiTrust is able to:

- Effectively support the goals of those libraries that operate under different copyright regimes than HathiTrust
- Address the the collaborative challenges of those libraries
- Provide strong support to all of our members
- Sufficiently engage them in our governance groups and processes

Membership is not the only aspect of our work internationally, and we pursue important work in collaboration with non-US libraries. As an example, in 2019 HathiTrust partnered with the British Library, the National Library of Scotland, and the National Library of Wales to explore how an international registry of digitized books could be created and sustained (see <https://gddnetwork.arts.gla.ac.uk/>). This project has helped us forge new alliances, to explore the challenges of metadata aggregation and normalization in a trans-Atlantic context, and to initiate discussions about “global” digital collection strategies.

- **Are there any proven strategies for demonstrating return on investment for our support and digital contributions to HathiTrust?**
- **How can we leverage the power of the HathiTrust collaborative to maximize benefit at our home institutions?**
- **How can HathiTrust facilitate collaboration between member libraries with similar collections and researchers with shared interests?**
- **Are there ways longstanding member institutions can support newer HathiTrust members for the good of all?**

During the 2019 Member Meeting, we spent time filling attendees in on HathiTrust’s work this past year towards meeting our [Strategic Directions, 2019-2023](#). Several questions were posed by registrants related to these so we wanted to include more information.

One of HathiTrust’s Strategic Themes is Empower where our goal is to empower our member libraries and their patrons through a durable, sustainable, and transparent organization and open infrastructure. Over the past year we took steps towards these goals by adopting new membership criteria and adjusting our fee model to support a



more equitable distribution of our costs to members. The nascent Member Toolkit will also better enable member libraries to promote HathiTrust on their campuses with greater consistency across the collaborative.

Additionally, we are looking for opportunities to increase connections between our members. Our goal is to facilitate collaboration between member libraries with similar collections and researchers with shared interests. Two outcomes we're working toward in 2020 include increased engagement and outreach work for both HTRC and HathiTrust and establishing the new PSC [User Engagement Task Force](#), which has identified addressing similar questions as part of their work.

HathiTrust has also framed a significant part of its Strategic Directions around the theme of Transformation, enabling our members to collectively address challenges that no single library can accomplish on their own. Our focus for the 2019 Member Meeting was the processes, programs, and activities that we can undertake to drive that transformation and increase the value of HathiTrust membership. During our conversations we spent some time hearing from speakers and attending to what change can mean from different perspectives. Taking this information forward, HathiTrust will start to address what change means internally and how it may impact our members.

Open Infrastructure and Scholarship

- **How does HathiTrust see itself in SPARC's Invest in Open Infrastructure Initiative announced earlier this summer? As libraries are asked to support more and more community led infrastructure projects, is there a way for HathiTrust to lead on demonstrating what sustainable practices and governance structures can actually look like?**
- **How is HathiTrust supporting the drive to a more open scholarly communications ecosystem?**

Over the last year or two we've seen the emergence of related, but independent efforts that are concerned with the scholarly communications infrastructure, a landscape of projects that are in various states of precarity for their future sustainability. [Invest in Open](#) has grown out of the ["2.5% Commitment"](#) described by David Lewis, and describes itself as "An effort to enable durable, scalable, and long lasting open scientific



HATHI TRUST

What You Want to Know Answers from HathiTrust Staff

and scholarly infrastructure to emerge, thrive, and deliver its benefits on a global scale.” Its initial work has focused on developing a [Census of Scholarly Communications](#) to help document the range of open source and open infrastructure projects and programs.

While Invest in Open takes a global approach, the Open Platform is very consciously located within the specific context of research libraries that have been and continue to fund these efforts, striving for “A coherent, interoperable, and adaptive ecosystem of systems and services under joint stewardship that collectively works to serve user needs and provide value to a diverse community of higher educational and cultural institutions.” **A not-so-subtle message of these projects might be stated as “We spend a lot of money on these startups — a more coordinated approach might produce a more strategically useful outcome.”**

HathiTrust’s financial and organizational strengths are directly relevant to these two community-based efforts, but our mission and programs are perhaps broader than their target. These two projects are heavily, though not exclusively, focused on open source digital preservation and access applications and services that are cooperatively developed by multiple partners. Some of these had extensive startup periods funded by foundations and funding agencies.

HathiTrust does not produce and distribute open source software or systems, and our organization’s startup was funded entirely by our founding and early members. However, we do believe that our collections and the associated metadata offer a significant platform for infrastructure development for shared print and collection management, text and data mining, and documentation of library holdings. See our [Strategic Directions, 2019-2023](#).

We aren’t immune from scrutiny by libraries examining their membership budgets, but we believe that we continue to offer a significant value to our members and the community at large. Our successes and support do offer some potential lessons in how to navigate a startup into maturity. Executive Director Mike Furlough serves as an advisor to the Census of Scholarly Communications project. HathiTrust has provided financial and organization data both to the Census and to Maurice York of the University of Michigan to support initial analysis for work on The Open Platform. We’ll stay involved in these discussions and advise them as called upon, because a strong, healthy scholarly communications ecosystem will reinforce the strength of our work.

December 5, 2019

Prepared by HathiTrust Staff



Collections: Diversity and Inclusion

- **I am interested in the topic of Hathi serving as a hub for unique research collections contributed by members and how members could assist in advancing that vision.**
- **Are there any initiatives in the works to make the HathiTrust collection more diverse and inclusive?**

HathiTrust's [Strategic Directions, 2019-2023](#) includes the need to “include many voices and perspectives in our collections” and enhance its quality, deepening the library with “a targeted, intentional strategy for new and retrospective publications.”

In 2020, we are launching a [Digital Collection Strategy Working Group](#) whose charge includes “...as its overall objective the continued diversification of the digital library’s holdings with regard to represented subjects, genres of holdings, years of publication, availability of in-copyright and out-of-copyright materials, language, and geographic origin.” The charge also includes these tasks:

1. Identify and encourage local initiatives by HathiTrust members to contribute new and underrepresented material to the corpus; support or leverage generalizable approaches that can be built upon by the HathiTrust member community.
2. Promote understanding of the relationship of HathiTrust holdings to other large-scale collections, with the specific goal of enhancing the diversity and strength of otherwise underrepresented digital materials in the HathiTrust Digital Library corpus.
3. Encourage member engagement in new material digitization and contribution to establish HathiTrust as a model for decentralized and collaborative collection design.
4. Highlight and support ways the corpus can build equitable access to the world’s knowledge, especially through the digital library’s support for accessibility and commitment to open sharing and distribution of the cultural and historical record.

This group will be getting underway in early 2020, and we will be drawing lessons from it continually.



We are also interested in learning more about what service expectations might be around collections, in order to better support discovery, presentation, and access at a collection level in HTDL. To make unique collections more readily available to researchers, we are undertaking a project in 2020 to improve how we share curated HTRC worksets that are available for computational analysis.

- **What can HathiTrust do to support a more inclusive academy? Creation of collections or data capsules highlighting on the history of underrepresented groups, or create fellowships and grants focused on creating a more diverse library/researcher workforce? The Colored Conventions Project at the University of Delaware and African American Digital Humanities at the Maryland Institute for Technology in the Humanities (MITH) are both models in this area.**

The Digital Collections Strategy Working Group is being formed and is charged to address the development of a more inclusive collection, and is expected to develop formal strategies to do this. We also encourage our members to be engaged in creating collections from the existing digital corpus that can highlight specific material for users (using the collection builder tool). We would love to have your contributions of digitized collections that bring underrepresented content and perspectives.

We can also consider strategies for more inclusive metadata and better ways to provide context within the digital library for materials that are potentially harmful, offensive and outdated.

Technical Infrastructure and Interoperability

- **What changes do you expect to interoperability of the HathiTrust infrastructure and system-to-system calling?**
- **What are HathiTrust's plans around IIIF API availability for institutions to integrate objects in the repository with locally developed layers (metadata, annotations)?**
- **Slow response time is an issue at various points in the research life cycle. What components can we collectively accelerate for researchers?**

HathiTrust continues to explore how to open up its data and services to a variety of uses. We are committed to providing our collected collaborative resources for the widest



possible legal use, in the most flexible ways for users at our member institutions. Our current service portfolio includes such interoperable services as metadata provision and full-text dataset provision for researchers. Currently we are exploring three aspects of a more open infrastructure:

1. HathiTrust developers have scoped a potential 2020 pilot project to test the implementation of IIIF protocols, starting with the Image API and the Presentation API. Many IIIF code pieces are in place and ready for this pilot, due to the extensive IIIF work of our development partners at the University of Michigan Library on other platforms implementing IIIF. We are interested in hearing from member libraries about desired use cases for IIIF interoperability.
2. As part of the near-term follow-up to the release in fall 2019 of the new [Policy on Metadata Sharing and Use](#), the new [Metadata Sharing Policy Task Force](#) will be reviewing use cases from members and vendors for consuming metadata of various kinds that are collected and managed at HathiTrust. Based on those use cases, HathiTrust staff, including our new Metadata Analyst, Graham Dethmers, will adapt some or all of our [existing bibliographic metadata provision methods](#), including possibly the HathiFiles, the Bibliographic Metadata API, and the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) feed. Staff may also respond to the new use cases by creating new APIs as needed, so that our bib metadata can be flexibly and selectively downloaded as needed into library catalogs, discovery systems, WorldCat, and research systems. We will assess use cases and possible solutions, including online help materials and reaching out to catalog managers at member institutions to support these changes.
3. HathiTrust Enterprise Architect Aaron Elkiss has begun his work this fall with a comprehensive review of the code involved in all HathiTrust applications and systems. He is considering whether restructuring modules and additional APIs might be appropriate for creating a more robust and flexible infrastructure. He is beginning to create a roadmap for re-design and development of the HathiTrust Digital Library applications over the next 2-3 years.

Access and Formats

- **Could the HathiTrust digital corpus be used to support controlled digital lending?**

December 5, 2019

Prepared by HathiTrust Staff



While we are very interested in the legal theories offered in support of controlled digital lending, we have not identified a way to implement it through HathiTrust that we believe to be lawful.

- **We would be curious to hear more about what HathiTrust is doing to change the print holdings information process (currently suspended for 2019 “in order to identify and implement workflow and infrastructure improvements”) and if/how this might affect the ability to reinstate a Section 108(c) access mechanism for partner libraries, among other things.**

Collection Services Librarian Natalie Fulkerson is leading a comprehensive assessment and redesign of [HathiTrust's print holdings process](#), which started in August 2019 with a broad stakeholder analysis and the setup of a cross-functional team. Since HathiTrust's print holdings database underpins multiple key collection-based services, this re-design must carefully respond to many needs. The team has begun planning the back-end framework for a new data management system that better integrates holdings records with other data stores including shared print retention commitments and collections analysis. The new system will also include better overlap reporting to members, more transparent calculations of fees, and easier incremental updates. Any proposed changes to the submission process or specification will be carefully vetted with member stakeholders prior to implementation.

A more flexible print holdings process will also help us to advance toward the goal of implementing online access to digitized versions of lost/missing and brittle/damaged materials that are held by U.S.-based member libraries, under the provisions of Section 108(c) of the U.S. Copyright Code. One of the several requirements for our provision of Section 108 access is that the material must be held by the library from which the user is attempting online access, with a holdings status of lost/missing and/or a condition status of brittle/damaged. Implementing Section 108 access for our U.S. members will also require us to establish a legally acceptable method of confirming access “from library premises” and to create a centrally coordinated database and process for accepting market availability reviews showing the materials to be out of print.



- **Currently HathiTrust doesn't accept born-digital materials. When do you plan to be able to bring in formats other than digitized print books, and what are the next steps toward that goal?**

We are continuing to work on the first pilot phase for ingesting born-digital materials into the HathiTrust Digital Library, with a dual focus: (a) adjusting our infrastructure — validation, ingest, display, and download — for accepting and managing EPUB files; and (b) resolving issues with accepting e-resources metadata for born-digital formats.

Because HathiTrust ingest processes were optimized early on specifically for acceptance of print book scans, with associated MARC records, the acceptance of born-digital formats forces us to adapt almost every part of our infrastructure. For the pilot, we are working with participants involved in a NEH-Mellon funded program, Humanities Open Books, including university presses and scholarly society publishers who are opening up selected monographs. We anticipate being able to provide access to these works in the first half of 2020, with bit-level preservation of the EPUB files. We will follow up with an assessment of downstream impacts, longer-term preservation possibilities, and other adjustments.

We will later turn our attention to the far greater range of challenges for ingesting PDF files. While the benefits of sharing and preserving PDF files are substantial, the wide range of PDF sub-formats, including some with multiple issues for us, will require us to proceed through carefully scoped stages. We will likely seek a small number of initial pilot partners who are willing to provide guidance and feedback as we tackle the associated issues sequentially with a series of pilots.

In the meantime, our Program Officer for Collections and Federal Documents, Heather Christenson, serves on the Depository Library Council's Digital Deposit Working Group that is exploring models for U.S. Government Publishing Office digital deposit. Her voice helps to ensure that HathiTrust is positioned to consider deposit of born-digital federal publications once this work comes to fruition.

- **How will HathiTrust expand access for member institutions to the entire shared corpus, including works not held in print by member institutions? How is membership to be valued with only partial access to the HathiTrust corpus?**



HATHI TRUST

What You Want to Know Answers from HathiTrust Staff

A common misperception among members is that member libraries have access to copyrighted items in the corpus. Particularly persistent is the misperception that members can access copyrighted items that they contributed or hold in their own collections. Under U.S. copyright law, this type of access is not permitted, except under special circumstances as specified in the Copyright Code, such as access provided to digital materials for print-disabled users. HathiTrust does, however, continue to work to expand lawful access to the corpus and make more robust the tools and access methods we provide.

Our [Commitment to Accessibility](#) has driven much of our work in this area in 2019. Access to the full corpus, including copyrighted materials, is available for students with print disabilities via the [Accessible Text Request Service](#) for designated disability service officers at any member institution. This service can now be extended to all HathiTrust member libraries that are located in a country that has signed and ratified the Marrakesh Treaty. In 2019, we upgraded the HathiTrust Digital Library applications to meet the needs of print-disabled users more closely, with extensive user testing and changes to come into conformance with the Voluntary Product Accessibility Template, including compliance with multiple international accessibility guidelines.

[The HathiTrust Research Center](#) now allows researchers to work on text and data mining with the entire corpus — even copyrighted items. The Extracted Feature dataset and HTRC algorithms for popular data mining actions have been expanded to include almost the entire corpus. Sharing of HTRC Data Capsules, where researchers have secure access to perform data mining activities, has been implemented for research teams. We are working to smooth the process for researchers to request and download datasets of open texts.

[Our Copyright Review Program](#) has accelerated the opening of works where copyright does not apply or whose copyright has expired, with improved training for the copyright reviewers from member institutions under the leadership of program manager Kristina Eden Hall, along with impressive productivity. We finished a project to open over 100,000 U.S. state government documents this fall. Kristina is planning a new copyright review project for 2020, focused on opening up Commonwealth government documents published in the UK, Canada, and Australia. She has put out a [call for participation in the Crown Copyright Project](#).

December 5, 2019

Prepared by HathiTrust Staff



HATHI TRUST

What You Want to Know Answers from HathiTrust Staff

Exploring additional lawful access is a focus in 2020 and we are currently working to determine whether we can lawfully open up certain front matter pages of copyrighted works, which would provide access to all users for additional helpful information about a book and its contents. Kristina Hall is working with developers from the Research Center on a machine learning project to develop an algorithm to computationally determine which pages are factual, non-creative pages in a text. We hope to analyze the success of that effort and its limitations in 2020.

Recognizing that access is a critical component of what our members seek, we continue to evaluate how we can expand legal access to as much of the corpus as possible for the widest range of qualified users. We are also engaged in discussions with members about how to go beyond HathiTrust Digital Library access to leverage collective action on collection management and growth.
