Digital Collection Strategy Working Group Background and Context

The Digital Collections Strategy Working Group supports the core mission of HathiTrust's collaborative agenda, which is the development and stewardship of the HathiTrust collection. In 2019, HathiTrust's Program Steering Committee charged the Digital Collection Strategy Working Group (DCSWG) to advance the strategic work of the Program Steering Committee and HathiTrust's operations around the digital collection.

This member-driven group was created to move forward the ongoing strategic focus on collection development as outlined in HathiTrust's 2019-2023 Strategic Directions. The plan states the need to "include many voices and perspectives in our collections" and enhance its quality, deepening the library "by a targeted, intentional strategy for new and retrospective publications" and by developing "new analytic capacities" for, among other uses, collection development.

Since its inception, DCSWG has focused on opportunities and approaches to bring more intentional collection development to HathiTrust Digital Library (HTDL) alongside already established mass digitization programs such as Google Books Project as well as opportunities that support and improve collection diversity across a broad range of factors such as language, region, underrepresented groups, or types of resources not appropriate for mass digitization or beyond the current scope of HTDL.

With this in mind, DCSWG has worked on designing approaches and methods to engage, analyze and understand what the value and purpose of HTDL is to HathiTrust members in the community and what methodologies and support systems would need to be developed and implemented to help members better understand and utilize the full potential of the collections in HathiTrust Digital Library.

With the onset of the COVID-19 pandemic in early 2020, DCSWG focused on opportunities and activities that suited virtual programming. The Working Group planned and hosted two discussion and member engagement sessions during Community Week 2020 that provided useful input to our planning. Over the course of 2021, DCSWG began working with Jessica Rohr to frame out a communications strategy that would extend our engagement with the HathiTrust membership. Planning included both outward communication to the members to help develop their knowledge of the HTDL collection and a survey that would gather further information from the members in support of DCSWG's work. Notably, in 2021 we also invested in the development of the HathiTrust Digital Collection Principles, with DCSWG member and Hathi staff member Heather Christenson leading this work. The Working Group was pleased to see the final Digital Collection Principles approved and published earlier this year.

In 2022, DCSWG completed and deployed the member survey, with active support from Jessica Rohr at HathiTrust. The survey results, summarized below, helped DCSWG understand the challenges that members face in engaging with HTDL, their local environment, and their needs

to participate in the shaping and growth of HTDL in support of both local and shared priorities. This input, together with that received in Community Week sessions, informed the development of the proposals that conclude this report.

Survey of individuals at HathiTrust member libraries: overview & summary of findings

The HathiTrust DCSWG proposed a survey as a mechanism to find out how to best communicate with the member libraries about the HathiTrust content, and how to communicate the HathiTrust digitization and ingest processes. The Working Group also hoped to gain insight on what resources, services, or information would be helpful to member libraries desiring to contribute (or contribute more) to increasing the diversity of the HathiTrust digital collection.

The survey was initially sent to multiple representatives at HathiTrust member libraries who identified as having collections responsibilities on April 8, 2022. A follow-up message to the HathiTrust member representatives list was sent on April 29. The survey was open for approximately one month, closing on May 13, 2022. The full survey is attached as Appendix 1.

We received 92 responses from 68 institutions, which represents 32% of the total HathiTrust member institutions. Of those institutions, 86% were in HathiTrust's Tier 2 or Tier 3 membership classification. See https://www.hathitrust.org/Cost for tier structure information. The individuals responding represented a wide variety of job titles, including: Library Dean/Director, Associate Dean, Assistant Dean, Collections Manager/Director, and other department heads or managers.

The survey was designed to learn more about each library's local landscape of collection development and digitization. For instance, we wanted to know if HathiTrust libraries have a strategic commitment to adding underrepresented material to their collections. We also asked about libraries' participation in Google Books digitization, and if the library was currently contributing to HathiTrust or to other digital libraries. For members who have not contributed to HathiTrust, the survey hoped to elicit information about what respondents thought would help to facilitate contributing their institution's digitized collections. The survey also asked if member libraries consider (or would consider) HathiTrust's stated collection priorities, in particular those expressed in the Digital Collection Principles, in their own collection and digitization policies and selection of materials.

The Working Group asked member libraries what information HathiTrust could provide that would be helpful. There were several aspects of this kind of information that had been relayed to the group anecdotally during our work, and the survey was a good opportunity to identify common lines of thought among member libraries. For instance, did libraries want HathiTrust to inform them of what titles or categories were needed? Do they want to know if HathiTrust has future plans for in-copyright works, particularly after the recent ETAS project. The survey also gave the opportunity for users to make other suggestions for approaches to diversifying the

HathiTrust digital collection.

The survey results provided insights into the importance of the HathiTrust corpus for members as well as factors that hindered more contributions. The Working Group found four recurring themes in responses: diversification of the corpus, lowering barriers to contribution, maximizing efficiencies, and contributing to a common good.

Helping to diversify the HathiTrust corpus and lowering barriers to contribution

Most respondents (78%) indicated that they include a commitment to adding underrepresented materials to their collection in their strategic plan, and a majority (55%) indicated that they would consider HathiTrust's overall collection priorities in local collection digitization decisions and contributions to the HathiTrust Digital Library (HTDL). The Working Group survey, however, found that technical requirements for deposit to HTDL and capacity for digitization were limiting factors for respondents, even at Tier 2 and Tier 3 membership levels. Additionally, there was interest in broadening what could be deposited to HTDL, including PDFs, ePUBs, and audio and video formats, as the types of files and content that HTDL can accept are sometimes not aligned with local needs or priorities. As one respondent commented:

> "Our current digitization efforts are mostly focused towards non-book special collections materials (newspapers, images, manuscripts). We would need lots of support to start digitizing monographs."

Note that this may also be an opportunity for further communication to HathiTrust members about the breadth of HTDL, since some non-book textual materials such as manuscripts may fit within HTDL's existing contribution scope. The types of resources currently accepted for HTDL are described here: https://www.hathitrust.org/ingest.

Another respondent remarked:

> "We are a well resourced library in terms of in-house programming expertise and even for us [the contribution process] created a very high bar for participation."

The commitment of adding underrepresented materials to collections, desire to participate in expanding the HT corpus, and interest in easier ways to contribute were strong themes in the survey responses.

Maximizing HT contributions and contributing to the common good

Members were very interested in knowing, in broad terms, what items their institution has that are not currently in the HTDL (40.3%), followed by wanting to know HathiTrust's plans for future access services for rights-protected materials (20.9%). Institutions wanted to be sure they could

efficiently contribute to HT with needed materials, helping to expand the corpus in the short-term while receiving guidance on future access to in-copyright titles. Additionally, the top three incentives for institutions to add digitized collections to the HTDL were: contributing to a public / common good (17.7%), preservation value of adding digitized content (16.8%), and increased discovery and access to collections (15.8%). While other options also ranked highly, the combination of members' desire to contribute to a common good with specific items needed in the corpus was an additional survey insight. Survey respondents noted:

> "I know HathiTrust is all about digitizing at scale, but now that you have so many millions of things added, it would help to add more of a fine-tuned approach."

> "[A primary factor for contribution is] knowing what content we have that HTDL would value."

> "From a teaching point of view, one of the most helpful things is the capacity to build collections within HT. It's great to have the entire collection available, but the narrower collections provide some needed focus, and often can help as substitutions for expensive commercial collections. "

The desire to ensure that the HathiTrust corpus is growing in needed ways, with a broader and more diverse collection for the common good, was an important takeaway in the survey.

Overall, the Working Group found that diversifying collections is a strategic priority for many HT member libraries, contributing to a common good is a strong incentive to add materials to HTDL, and institutions that can add diverse materials can help advance the goals for all member libraries. To these ends, survey respondents identified selection guidance on materials to digitize, additional support for digitization and ingest, and widening accepted digital formats as ways to achieve these strategic priorities with HathiTrust.

Proposals to PSC for projects, services, or resources to support HTDL collection development

The survey responses, combined with information from two Community Week events in 2020, provided good input to the Working Group's discussions and planning. Pursuing more refined, but collaborative, development of the HathiTrust Digital Library will require participation from across the HathiTrust membership. With this in mind, the Working Group felt that it was important to hear directly from representatives of the member libraries what they believed would support their institution's participation in this shared goal, and the survey helped to elicit this information. We also hoped to provide opportunities for further member engagement in that work and approaches that would be accessible across the broad range of HathiTrust members. In framing the proposals that follow, the Working Group strove to identify projects, services, or activities that incorporated these goals, bringing a variety of approaches and varying resource

requirements to development of the HTDL.

As a working group charged by the Program Steering Committee, DCSWG brings these proposals to PSC as possible approaches to further development of HTDL. We recognize that they will need additional discussion and review by PSC and with HathiTrust staff and the Board of Governors, and that they may be endorsed or enacted only selectively. We anticipate that PSC will provide DCSWG with further guidance or direction as necessary to align the Working Group's activities with the goals and priorities of HathiTrust.

Proposal 1: PSC work with HathiTrust staff to establish a HathiTrust staff-coordinated group of member volunteers (possibly a User Support subgroup with a focus on the process of contribution) to assist member libraries with content deposit. Explore and document for future stages of work what changes would be necessary to lower technical barriers to contribution.
Description (context, etc.): DCSWG investigations have consistently found that member libraries considering content contribution are deterred by the technical requirements (file formats, metadata) of the ingest workflow. Enabling more member libraries to contribute content will have the primary collections benefit of expanding and diversifying HTDL content, as well as the secondary benefit of making available to more member libraries the long-term digital preservation benefits HathiTrust offers. The member volunteer group approach will also create a new opportunity for employees of member libraries to become more engaged with HathiTrust. Such a group should work closely with HathiTrust metadata and collections staff to ensure understanding and coordination of the process of contributing content.
Resources needed: HT staff time to manage coordination of volunteer activities. Member volunteer time.
Limitations or caveats: This proposal addresses only the technical barriers to content contribution, not the need for prioritization of content to contribute (see Proposal 3). An additional, related barrier is member libraries' limited staff and equipment capacity for digitization, which would not be mitigated by this proposal.

Direct support for digitization at the member libraries, either in funding or by providing or coordinating digitization services, would be a strategic change in the current scope of HathiTrust's activities. Survey responses indicate that portions of the membership felt this would be valuable, however, and additional investigation may be worthwhile. A decision to move in this direction would require careful consideration from HathiTrust staff, PSC and the Board of Governors, and the membership, as it would have significant resource, staffing, and cost implications.
Proposal 2: Build a requirements document for infrastructure and staffing needs to expand the file formats that can be ingested into HTDL to support resource planning by HathiTrust, PSC, and the Board of Governors for implementation.
Description (context, etc.): Allowing more flexibility in acceptable formats would support diversifying the HTDL by opening it to new categories of content, such as born-digital ETDs and

government documents. The Working Group was also reminded in our consultations with members that PDFs are ubiquitous in member libraries' local repositories. While these are not generally considered "archival quality" files, they are not transitory in the members' digital collections and preservation of these materials is a concern. Adapting our practices to support ingest to HTDL would not only expand HTDL content but also allow members to consider and address preservation needs for these digital resources at scale. Audio and video formats, while further from Hathi's tradition of textual materials, could diversify the collection with content such as oral histories or locally-produced media.

Resources needed: Assessing the impacts to infrastructure and ongoing digital asset management and access of expanding the acceptable HTDL formats will require a significant investment of time from HathiTrust staff. Implementation would require additional funding and staffing to develop processes for varying file formats and appropriate metadata and access methods. Audio and video files, if included, are storage-intensive resources, and infrastructure requirements would need to be planned and funded accordingly.

Limitations or caveats: Identifies needed resources but doesn't address immediate needs that member libraries may have.

Proposal 3: PSC appoint a member volunteer task force (or modify DCSWG's charge) to develop structured criteria, guidelines, or models to assist member libraries in selecting and prioritizing materials for digitization and contribution to HTDL

Description (context, etc.): DCSWG investigations have consistently found that member libraries would like more guidance about prioritizing materials for digitization and ingest, and a willingness by members to consider HathiTrust goals in their selection of materials for digitization. Guidance for both mass digitization program participants (e.g., Google Books) to get the most from these programs and for local digitization projects would be beneficial. Documentation that provides broad guidelines or models to support selection could be developed and refined with member feedback by such a group.

Respondents to the DCSWG survey also indicated an interest in lists of their physical holdings that are not represented in HTDL. General guidance for working with the existing HathiTrust holdings report format to produce lists could be developed by such a group, but additional analysis and reporting of member collections and overlap, an area that HathiTrust is exploring, is likely to be beneficial.

Resources needed: Member volunteers, with participation from HathiTrust staff. A deeper analysis of member libraries' collections and gaps/overlap across the HathiTrust membership would require additional collection analysis expertise and tools. Developing new analysis and reporting capacity may require additional staffing.

Limitations or caveats: This proposal does not address technical barriers to content contribution or grow local capacity to digitize physical collections.

Proposal 4: PSC appoint a member volunteer task force to assess special uses of HTDL content (e.g., HTRC, ATRS) as an approach to identify collection gaps and consider a process for adding content 'on-demand' to support these services.

Description (context, etc.): HTRC and ATRS support users with needs and access requirements that differ from other Hathi users. Both areas may also be more sensitive to timing than standard digitization and deposit paths. Broadening content available to ATRS, in particular, supports HathiTrust's goal "To dramatically improve access to these materials in ways that, first and foremost, meet the needs of the co-owning institutions, with a particular emphasis on ensuring access for individuals with print disabilities."

Resources needed: Member volunteers, with participation from HathiTrust staff. Support from HathiTrust governance (PSC, Board). Assessment of content used through these services may require collection analysis expertise and tools. Developing new analysis and reporting capacity may require additional staffing.

Limitations or caveats: HTRC and ATRS are specialized services that may currently have lower use than other areas of the HTDL, but the potential impact for users of these services is significant.

Appendix 1 - Survey instrument

The HathiTrust Digital Collection Strategy Working Group is seeking your input to inform the next stage of its work. The survey below consists of 10 questions and should take no more than 10-15 minutes to complete.

The Working Group is charged, in part, with identifying possible approaches to improving collection diversity (e.g., language, region, underrepresented groups) in the HathiTrust Digital Library (HTDL) and with helping HathiTrust members to gain a better understanding of the HTDL overall. Your responses to the questions below will assist the Working Group in identifying areas where additional communications to the members about HTDL content or processes would be helpful. Importantly, your responses will also provide key information on which services, supports, or programs individuals at HathiTrust member libraries believe would most significantly advance the work of bringing greater diversity to the HathiTrust Digital Library.

The Working Group appreciates your time and assistance.

1. To help the Digital Collection Strategy Working Group contextualize your responses, please tell us your:

   Institution:
   Position:

E-mail:

2. Does your library's strategic plan include a commitment to adding underrepresented
   materials to your collection?
   Yes
   No
   I don't know

3. Does your institution contribute digitized content to the HathiTrust Digital Library
   (HTDL)?
   Yes
   No
   No, but expecting to start within the next 12 months
   I don't know

4. Is your library a GoogleBooks partner?
   Yes
   No
   I don't know

5. Does your institution regularly contribute to another digital library? If so, which one(s)?
   Internet Archive
   DPLA
   JSTOR Community Collections
   Other (please specify) + text box (short, with a character limit)
   No, but planning to start within the next 12 months
   No, with no plans to begin
   I don't know

6. What factors would facilitate your contribution of digitized content to the HathiTrust
   Digital Library (HTDL)?
   Better understanding the potential value of depositing
   Talking points to support conversations with leadership at my institution on the value of
   contributing to HTDL
   More information about how HathiTrust libraries participate in mass digitization
   programs
   A guide to the process of depositing digitized collections in HTDL
   A stronger understanding of how copyright law plays a role in use and access in

HathiTrust

More support from HathiTrust for contribution (e.g., metadata support, technical support)

A cohort of peers that supports knowledge sharing

Other (please specify) + text box (short, with a character limit)

7. What in your view are/would be incentives for your institution to add digitized collections to HathiTrust Digital Library (HTDL)?

Preparation for current and future needs for emergency access

Diversification of the existing HTDL collection

Contributing to a public/common good

Preservation value of adding digitized content to HTDL

Increased discovery and access to your library's collections

Institutional reputation

Support for machine learning methods via HathiTrust Research Center

Other (please specify) + text box (short, with a character limit)

8. Would your institution be willing to consider HathiTrust's overall collection priorities, as expressed in the [Digital Collection Principles,](#) in your local collection digitization decisions and contributions to the HathiTrust Digital Library?

Yes

No

I don't know

9. In thinking about your own library's priorities, what would be helpful for you to know about HathiTrust or the HathiTrust Digital Library? (please rank your choices)

Knowing in broad terms what HathiTrust would like added to the collection (i.e., HathiTrust collection development priorities)

Knowing in broad terms what HathiTrust end users are using or requesting (i.e., are there categories of material that users at HathiTrust libraries ask for)

Categories of material that would fill gaps in the HathiTrust Digital Library (e.g., books in non-English languages, books authored by underrepresented groups, resources covering specific topics or regions)

A list of items my institution holds that are not in HathiTrust Digital Library

HathiTrust's plans for future access services for rights-protected materials, such as in-copyright titles

Other (please specify) + text box (short, with a character limit)

10. Please select up to 3 potential area(s) of activity that you believe would most improve

collection diversity (e.g., language, region, underrepresented groups) in the HathiTrust Digital Library (HTDL).

In making your selections, assume that all necessary resources are available.

I would like HathiTrust to:

Provide more analytics of members' collections to assist members in targeting for digitization materials not in the HathiTrust Digital Library (HTDL)

Create a mechanism for member libraries without robust digitization programs to contribute content to HTDL (e.g., partnership/sponsorships between member libraries, coordination with Internet Archive, direct support for digitization from HT) Support deposit of digitized collections from non-member libraries

Subsidize the recruitment of new members that could bring a diversity of collections to HTDL

Establish a HathiTrust program (e.g., similar to the Shared Print Program) to coordinate collaborative HTDL development among participants

Develop support within HathiTrust to facilitate deposit of material digitized as part of grant-funded projects at member libraries (e.g., craft standard language for grant applications, develop a toolkit to support deposit of materials)

Actively solicit contributions of known large sets of mass digitized content that are not already in HathiTrust

If you have questions, please contact feedback@issues.hathitrust.org