



HathiTrust Overview

Contents

- [Partnership](#)
- [Introduction](#)
- [Governance](#)
- [Cost](#)
- [Services](#)
- [Preservation](#)
- [Access](#)
- [Shibboleth](#)
- [Collection Builder](#)
- [Rights Management](#)
- [Development Environment](#)
- [Current and Future Work](#)

Partnership

Introduction

HathiTrust is an international partnership of more than [seventy research institutions](#) and libraries that are working together to ensure the long-term preservation and accessibility of the cultural record. The partnership launched a digital repository in 2008 that currently contains over 10.5 million volumes, digitized from the partnering library collections. More than 3.2 million of these volumes are in the public domain and freely viewable on the Web. The partners aim to build a comprehensive archive of published literature from around the world and develop shared strategies for managing and developing their digital and print holdings in a collaborative way. The primary community that HathiTrust serves are the members (faculty, students, and users) of its partners libraries, but the materials in HathiTrust are available to all to the extent permitted by law and contracts, providing the published record as a public good to users around the world.

Information about the benefits of partnership is available at http://www.hathitrust.org/features_benefits.

Governance

HathiTrust is governed by a [Board of Governors](#), established in April 2012 ([view information about Board elections and composition](#)), which manages the budget and finances of HathiTrust and makes decisions on all aspects of the partnership. The daily activities of HathiTrust are managed by an Executive Director.



HathiTrust accomplishes its work through the formation of working groups, which are composed of members from throughout the partnership. Participation in working groups is the primary mode of participation in governance for new partners joining HathiTrust. More information about working groups is available at http://www.hathitrust.org/working_groups.

Cost

Partners pay for the basic infrastructure costs of content preserved in HathiTrust, which include the costs of storage, backup, data centers, servers, and some staff. The model is described in full [here](#). A [FAQ](#) on the model is also available.

Services

Preservation

HathiTrust offers long-term preservation (including bit-level preservation and migration) of deposited materials. This is accomplished through the inclusion of standard and open content formats that we are confident can be preserved and migrated forward to new formats over time, rigorous validation of content on ingest, reliance on standards for metadata such as METS and PREMIS, reliance on standards for repository design and trustworthiness such as OAIS and TRAC, and regular checks on the integrity of stored content (through automated system checks, user access, and repository processes such as full-text indexing that use the content on a regular basis). HathiTrust has two active storage sites, one in Michigan and one in Indiana, and a copy of all content is stored on backup tape in a third location in Michigan. Storage is replaced every 3-4 years.

Content formats currently supported are ITU G4 (bitonal) TIFF, JPEG2000, JPG, and Unicode OCR with and without coordinates. A full description of policies and specifications for submitted content can be found at the following URLs:

[HathiTrust Guidelines for Digital Object Deposit](#)
[HathiTrust Deposit Form](#)

Access

HathiTrust is a “light” archive and the partners strive to provide as much access to content in the digital library as legally possible. This includes viewing, downloading, and searching access to public domain volumes, and searching access to in copyright volumes. Specialized features are also available which facilitate access by persons with print disabilities, and allow users to gather subsets of the digital library into “collections” that can be searched and browsed. Partners may offer print-on-demand services for public domain volumes they have digitized via HathiTrust (partner institutions develop their own print-on-demand agreements with entities such as Amazon).



Providing access in the context of copyright laws and third-party restrictions on content (such as those imposed by Google for content it digitizes), results in a slightly complex landscape of access and use scenarios. These are described in the table and accompanying notes below. There are four basic types of works in HathiTrust for access purposes, each of which is available based on the audience that is accessing them and/or the applicable laws and restrictions.

Type of work	Searchable (bibliographic and full-text)	Viewable	Full-PDF download	Print on Demand [1]	Print disabilities [2]	Preservation uses (Section 108) [3]
Works that are public domain when viewed from anywhere in the world. These are works published anywhere in the world prior to 1873.	Worldwide	Worldwide	Worldwide if not scanned by Google. If scanned by Google, only available to partners.	Worldwide	Partners worldwide (see note 2)	N/A
Works that are public domain only when viewed from the United States (see http://www.hathitrust.org/copyright). These are works published outside the U.S. between 1873 and 1923.	Worldwide	When accessed from within the United States[4]	When accessed from the United States if not scanned by Google. If scanned by Google, only available to partners within the U.S.	Available within the United States	Partners in the U.S. (worldwide where similar laws are in effect) (see note 2)	N/A
Works that rights holders have opened access to in HathiTrust	Worldwide	Worldwide	Worldwide (if digitized by Google, full download only available with CC license)	Worldwide with permission	Partners worldwide (see note 2)	N/A
Works that are in-copyright (this includes works whose copyright status is undetermined)	Worldwide	Not available	Not available	Not available	Partners in the U.S. - worldwide where similar laws are in effect (see note 2)	Partners in the U.S. where applicable (see note 3) - worldwide where similar laws are in effect



- 1 We hope to offer print-on-demand services for all public domain materials in HathiTrust. Arrangements for print-on-demand are made on an institutional basis. The University of Michigan has negotiated agreements with Hewlett Packard to prepare images for printing and with Amazon to distribute the printed volumes on demand. It is possible for joining institutions to add themselves to Michigan's existing agreements with HP and Amazon or to conclude their own agreements with the entity or entities of their choice. The University of Michigan and the University of California are the only institutions currently offering print-on-demand services for materials.
- 2 HathiTrust has created a customized interface that is available worldwide to make texts that are viewable in HathiTrust (those that are in the public domain or open access) more accessible to users with print disabilities. Under provisions of U.S. copyright law, the University of Michigan has taken this a step further and is delivering the full-text of any work that is or has been owned by the University of Michigan Library, whether public domain or in copyright, digitized by Google or not, to users who are registered with the University's Office of Services for Students with Disabilities. HathiTrust is working on expanding this kind of access to other partner institutions, including those outside the United States where allowed by law. The technical infrastructure for expanding access is fully in place (a policy framework for certifying users as having a disability must be in place at the individual institutions), but release of the service has been delayed due to concerns about starting a new service for accessing in-copyright materials when the Authors Guild lawsuit is getting underway. The timeline for release is still to be determined, but it is not tied to the lawsuit. Partners will need to release a particular shibboleth attribute for their users to gain access. See information about [Shibboleth in HathiTrust](#).
- 3 Section 108 of U.S. copyright law allows libraries in the U.S. to make a digital copy of works that they hold or have held in their print collections available to users on library premises if the print work is lost, damaged, deteriorating, stolen, and not available at a fair market price. This service, like augmented access for users with print disabilities, is currently fully implemented only at the University of Michigan, where users are able to read digital copies of qualifying volumes. We are able to provide print replacement services to other partner institutions currently under Section 108 or under applicable laws for institutions outside the United States, but are not yet providing reading access to digital volumes. If you are interested in helping us to record the necessary information about volumes in HathiTrust to provide this service (e.g., in the United States, that they are not available on the market at a fair price), please let us know.
- 4 We do not have resources currently to analyze the status of works in countries outside the U.S. (though significant work is being done to determine the status of "orphan" works published within the U.S.), so this an area of future work for us. For institutions outside the U.S. that become partners, a possible way to move forward would be to incorporate the laws of those countries into our rights determination process, and use login via Shibboleth as a way to provide access to volumes that are in the public domain in those countries, to the users in those countries.

Shibboleth

HathiTrust uses [Shibboleth](#), a mechanism for inter-institutional authentication, as a means to provide specialized services in HathiTrust to persons affiliated with partner institutions. These



services currently include full-PDF download for public domain works and facilitated access to the Collection Builder, which makes it possible for users to aggregate works into permanent collections either for private use or to share publicly with others. Shibboleth will also be the foundation on which HathiTrust is able to deliver services for users with print disabilities and access to orphan works.

Collection Builder

In addition to bibliographic and full-text search, a [Collection Builder](#) tool exists to facilitate use and discovery of HathiTrust materials. Items can be added to permanent public or private collections from the PageTurner application (e.g., <http://hdl.handle.net/2027/mdp.39015079129154>), or from the full-text search results page (e.g., <http://babel.hathitrust.org/cgi/lis?q1=Eyjafjallajokull&a=srchls>). Individuals at partner institutions that are configured with Shibboleth may login to Collection Builder through their institutional login. Others may login after creating a University of Michigan [Friend Account](#).

Rights Management

HathiTrust performs rights determinations on all content that enters the repository, using fields in bibliographic metadata (which must be in MARC format) such as publication year and location, and author death date (see the [description of the rights determination process](#)). This information is stored in a rights database, which is described at http://www.hathitrust.org/rights_database. Further information about rights is available in note 4 in the Access section above and at <http://www.hathitrust.org/copyright>.

Development Environment

HathiTrust has configured a collaborative environment to facilitate distributed development of HathiTrust applications by partner institutions. Partners that are interested in engaging in development or exploring the environment should contact us at feedback@issues.hathitrust.org.

Current and Future Work

A list of the working groups that currently active in HathiTrust can be found at http://www.hathitrust.org/working_groups, and grant proposals and projects at <http://www.hathitrust.org/projects>. Some primary areas of work for us right now are:

- 1 Scaling up ingest of partner content that was not digitized by Google or the Internet Archive. We have efficient mechanisms for ingesting content from these sources but are expanding to accommodate the variety of content owned by our partner institutions.
- 2 Adding new content types. We are currently involved in pilot projects around ingest of image and audio content, and born-digital publications.
- 3 Assessing the quality of content in the repository. A grant project HathiTrust is participating in is described at <http://www.hathitrust.org/projects>.



HATHI TRUST

- 4 Expanding manual copyright review of works. The University of Michigan recently received a second IMLS grant to expand the review work done under the IMLS-funded [Copyright Review Management System](#) to non-U.S. publications.
- 5 Doing more formal work around collections. HathiTrust charged a [Collections Committee](#) in July 2010 that is working on a variety of issues related to collections in the repository (see the [Collections Committee charge](#)).
- 6 Communications, Usability, User Support. HathiTrust has also charged working groups in areas of communications (outreach to particular audiences such as faculty and students, and to current and potential partners), usability of HathiTrust tools and services, and user support.
- 7 Building the [HathiTrust Research Center](#).
- 8 Using HathiTrust as a backend for publishing born-digital materials.
- 9 Transitioning to new governance structure and engaging new initiatives decided on at the [2011 Constitutional Convention](#).
 - This includes initiatives geared toward enhancing access to US federal government documents and the establishment of a distributed print monograph archive.