



národní  
úložiště  
šedé  
literatury

## **HathiTrust and collaborative collection services**

Weltin, Heather  
2019

Dostupný z <http://www.nusl.cz/ntk/nusl-407856>

Dílo je chráněno podle autorského zákona č. 121/2000 Sb.

Licence Creative Commons Uveďte původ-Neužívejte komerčně-Nezpracovávejte 4.0

Tento dokument byl stažen z Národního úložiště šedé literatury (NUŠL).

Datum stažení: 05.05.2024

Další dokumenty můžete najít prostřednictvím vyhledávacího rozhraní [nusl.cz](http://nusl.cz) .

---

# HathiTrust Collections & Collaborative Collection Services

---

16th IFLA ILDS Conference

Heather Weltin, HathiTrust Shared Print Program Officer

October 11, 2019



# Today's Plan

---

## 1. HathiTrust

- History
- Mission and Vision
- Membership

## 2. Digital Collection

## 3. Collaborative Programs

## 4. Access and Use

... or Beyond the Paywall in the HathiTrust Realm



# History

ENGLISH ESPAÑOL 中文

## The New York Times

ARCHIVES | 2004

### Google Is Adding Major Libraries to Its Databases

By JOHN MARKOFF and EDWARD WYATT DEC. 14, 2004

Google, the operator of the world's most popular Internet search service, plans to announce an agreement today with some of the nation's leading research libraries and Oxford University to begin converting their holding into digital files that would be freely searchable over the Web.

It may be only a step on a long road toward the long-predicted global virtual library. But the collaboration of Google and research institutions that also include Harvard, the University of Michigan, Stanford and the New York Public Library is a major stride in an ambitious Internet effort by various parties. The goal is to expand the Web beyond its current valuable, if eclectic, body of material and create a digital card catalog and searchable library for the world's books, scholarly papers and special collections.

### An Elephant Backs Up Google's Library

BY MIGUEL HELFT OCTOBER 13, 2008 6:17 PM

Google often says that it likes to take the long-term view of things. But Google's idea of long term does not appear to be long enough for some librarians, who tend to equate long term with forever, at least when it comes to preserving books.

On Monday, a group of major libraries that are participating in Google's Library Project, said they are working together to create what amounts to a publicly accessible backup of the digital library that Google is creating. The project, [which is called HathiTrust](#), includes libraries at 12 Midwestern universities like the University of Michigan, the University of Iowa and the University of Illinois, plus the 11 libraries of the University of California system. (Hathi is Hindi for "elephant," an animal that is s

In the [Google Books Library Project](#), the scanning the collections of several large gives users access to the complete text of public domain, and to snippets of books copyrights. Google also gives each library digitized from that library.



October 13, 2008

HATHI TRUST

*HathiTrust launches with 12 founding members from the Committee on Institutional Cooperation and the 11 libraries of the University of California.*



# Mission and Vision

---

## Mission

The mission of HathiTrust is to contribute to research, scholarship, and the common good by collaboratively collecting, organizing, preserving, communicating, and sharing the record of human knowledge.

## Vision

HathiTrust will be a vital catalyst for emerging forms of research, teaching, and learning that engage the scholarly and cultural record.



# Membership

---

## Eligibility

Academic and research institutions worldwide.

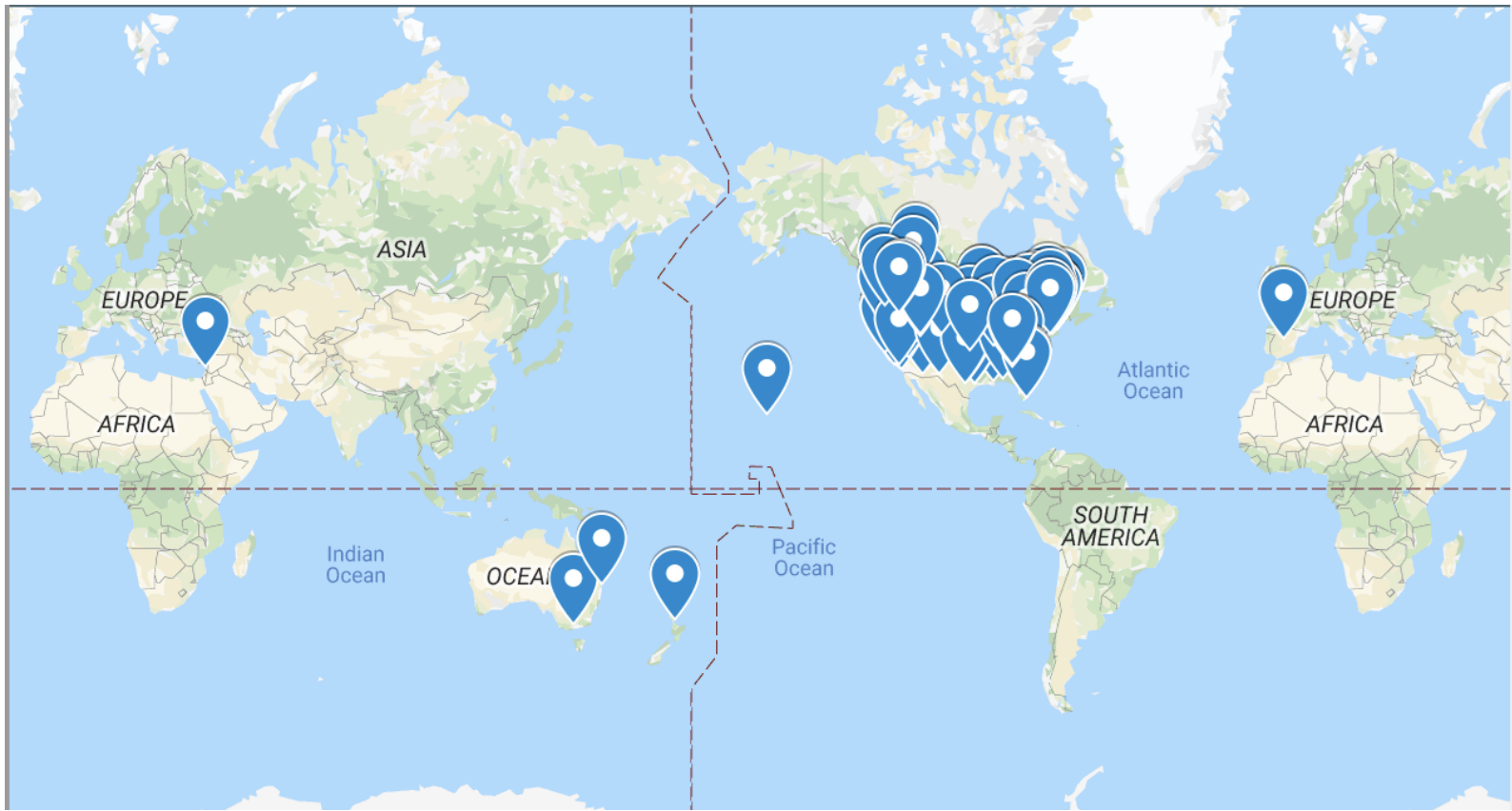
## Requirements

- Required
  - Submission of holdings data on physical collections.
  - Implementation of SAML-based authentication system for access services.
  - Annual fees.
- Not required, but encouraged
  - Deposit of collections
  - Participations in working groups, governance



# Membership Across the Globe

---



# HathiTrust Infrastructure

---



HATHI  
TRUST



HATHI TRUST  
research center



UNIVERSITY OF  
MICHIGAN



ILLINOIS  
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN



University of California  
**CDL**  
California Digital Library



INDIANA UNIVERSITY





---

# HathiTrust Collection

---



# Collections - Currently Digitized

---

- 16,987,842 total volumes
- 8,237,760 book titles
- 454,195 serial titles
- 5,945,744,700 pages
- 761 terabytes
- 201 miles
- 13,803 tons
- 6,460,113 volumes(~38% of total) in the public domain



# Top 10 Contributors to HathiTrust, July 2019

Library	Deposited Volumes
University of California Libraries	235,605
University of Illinois	97,793
University of Michigan	56,468
Ohio State	54,688
Michigan State University	43,334
University of Minnesota	22,365
Indiana University	14,402
University of Virginia	13,012
Cornell	10,258
University of Texas	7,807

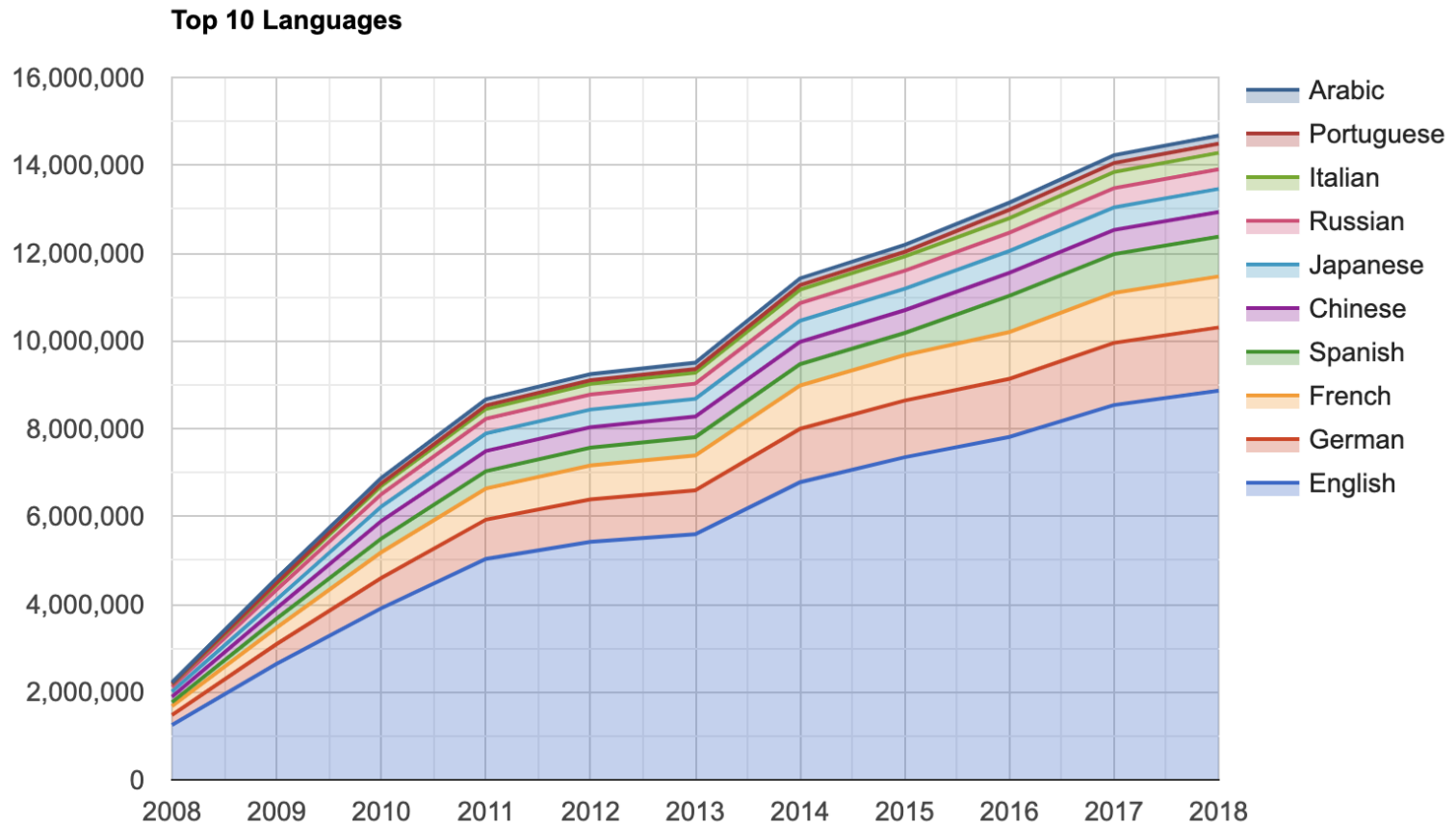


# Collections by Publication Date

Publication Dates as of October 2018



# Collections by Language (Top Ten)



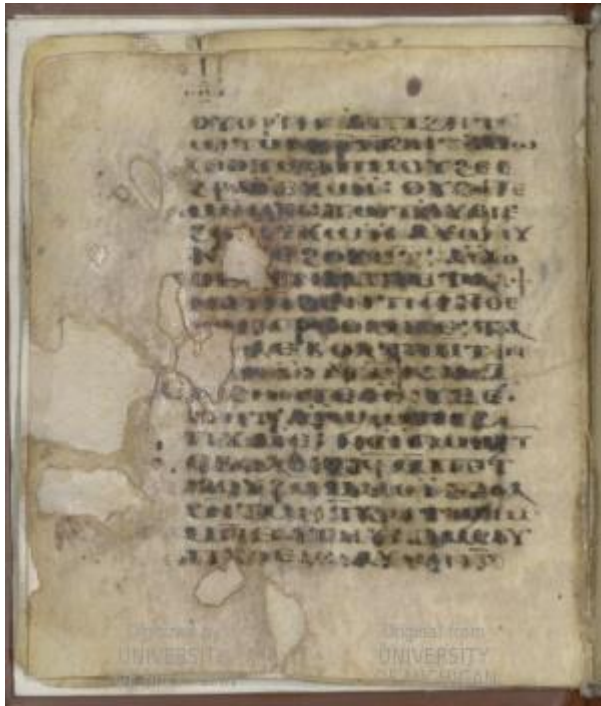
Class	Description	# of Records
A	GENERAL WORKS	52,925
B	PHILOSOPHY. PSYCHOLOGY. RELIGION	333,191
C	AUXILIARY SCIENCES OF HISTORY	59,964
D	WORLD HISTORY AND HISTORY OF EUROPE, ASIA, AFRICA, AUSTRALIA, NEW ZEALAND, ETC.	540,697
E	HISTORY OF THE AMERICAS	96,330
F	HISTORY OF THE AMERICAS	278,174
G	GEOGRAPHY. ANTHROPOLOGY. RECREATION	111,267
H	SOCIAL SCIENCES	613,535
J	POLITICAL SCIENCE	141,800
K	LAW	108,474
L	EDUCATION	120,108
M	MUSIC AND BOOKS ON MUSIC	115,180
N	FINE ARTS	165,334
P	LANGUAGE AND LITERATURE	1,105,730
Q	SCIENCE	314,816
R	MEDICINE	129,411
S	AGRICULTURE	113,591
T	TECHNOLOGY	217,405
U	MILITARY SCIENCE	26,615
V	NAVAL SCIENCE	10,330
Z	BIBLIOGRAPHY. LIBRARY SCIENCE. INFORMATION RESOURCES (GENERAL)	94,230

# Subjects of HathiTrust Content - as derived from Library of Congress Classification classes

Note: LCC classes are only present in 60% of HathiTrust records.



# Collection Images



[Psalter]. , 6.

## #exstrange: CURATORIAL AND ARTISTIC INTERVENTIONS ON EBAY

MARIALAURA  
GHIDINI &  
REBEKAH  
MODRAK

Marialaura Ghidini and Rebekah Modrak are the lead curators of *exstrange*. This essay introduces the project, its structure and curatorial parameters, and its use of the online marketplace as exhibition platform.

This text was written for the *Journal of Curatorial Studies* and was published in Volume 6, Number 1, April 2017. It has been slightly edited here to introduce a wider example of artworks produced during *exstrange*.

#exstrange was an online curatorial project generating a collection of artworks conceived by contemporary artists and designers to be encountered, auction-style, by the users of one of the largest marketplaces on the web, eBay. The artists creating artworks on auctions on *exstrange* used the eBay interface and listing template as the tool of production for their work. As such, the chosen sale category—from Business & Industry and Consumer Electronics to Tickets & Experiences, for example—as well as title, descriptive text, accompanying images and pricing, all constituted the work. The interface of the e-commerce site became the space in which the artworks resided and were interpreted in consonance with the specificities determined by their design, notably the time-limited, one-to-one user engagement and socio-linguistic interactions pertaining to online commerce. Thus, *exstrange* emerged as an auction created on an ongoing configuration among the one billion items for sale across the various national eBay platforms.

Launched on 15 January 2017, *exstrange* presented an artwork on auction a day until 8 April 2017. Directed by the tag *exstrange* in the listing title, the works could be found in that vast archive of commodities online that is eBay. We began the project by inviting 31 artists whose practices resonated with the project. Some of them were concerned with issues related to digital

production and creative labor (such as Silvio Lorusso, Geraldine Abbott, Tyler Demmead, and Maria Micaela and Marie Neumann), some with the commodification of objects and social dynamics (such as Fendling, KOOSE and Lloyd Corporation), while others with processes of technological historicization and consumer culture (such as Alireza Hosseini, Anshara Hande and Tara Kulkarni). We also asked 17 curators living in different parts of the world to each invite three artists according to their own interpretation of our curatorial invitation (see 1-6). The guest curators wrote a statement accompanying their interventions and, in some instances, created their own tags to identify them; for instance Hana O' Murcha adopted #nothingisnewer while Gela Tardone used #veryhandmade. Our intention was to open up the project to perspectives with diverse critical literatures but also heterogeneous social and geographical contexts—setting up a process that, along with the project's open call to participation ("Artist-Jurors"), would go beyond our curatorial control.

Rather than initiating a new web platform, we sought to work with the properties and functionalities of an already existing online service, one governed by the logic and language of commerce. eBay was founded in 1996 by the Iranian-American entrepreneur Pierre Omidyar. Its tacit anthropocentric model of person-to-person commerce through time, if proposed

UNIVERSITY  
OF MICHIGAN

Ghidini, Marialaura, Rebekah Modrak, and Michigan Publishing (University of Michigan). #exstrange: a Curatorial Intervention On Ebay. Ann Arbor, Michigan: Michigan Publishing,



---

# Collaborative Programs

---





# Collaborative Programs - User Support

---

In order to efficiently and effectively serve the needs of users, and ensure the highest level of function of HathiTrust services, HathiTrust has charged a multi-institution user support team.



# Collaborative Programs - Copyright Review Program

---

The Copyright Review Management System (CRMS) has enabled large-scale, distributed investigations into the copyright status of works in the HathiTrust collection.



# Collaborative Programs - U.S. Federal Documents Program

---

## Mission:

Through coordinated and collective action, expand and enhance digital access to U.S. federal publications including those issued by GPO and other federal agencies.

As of July 1, 2019

- 542,731 bibliographic records
- 1,353,343 digital objects
- 509,600 monographs
- 32,619 serial titles



# Collaborative Programs - Shared Print

---

## Goal

The primary goals of the HathiTrust Shared Print Program are to ensure preservation of print and digital collections by linking the two, to reduce overall costs of collection management for HathiTrust members, and to catalyze national/continental collective management of collections.

As of May 2019:

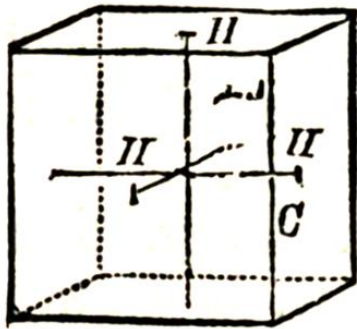
- 79 retention libraries participating in HathiTrust Shared Print Program
- 17,889,303 single-part monographs
- 5,476,598 individual titles



# Collaborative Programs - HathiTrust Research Center

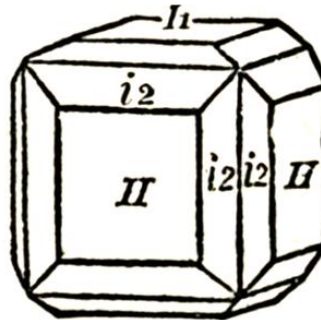
---

## Featured Services



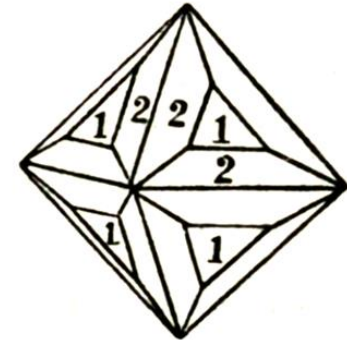
### Extracted Features

An unrestricted dataset of metadata and word counts for each page in the HathiTrust Digital Library. Download and explore on your own machine.



### Text Analysis Algorithms

Web-based, click-and-run tools that perform computational text analysis on worksets, which are user-created collections of volumes. No programming required.



### Data Capsules

Secure virtual environments for non-consumptive text analysis, where researchers can implement their own data analysis and visualization tools.



# Collaborative Programs - Projects and Grants

---

## **Global Registry of Digitized Texts Project**

Project to develop an international registry of digitized texts enabling discovery and use.

## **Federated Module for Accessible Material for Education (FRAME)/Accessibility Grant**

Project looking at facilitating access to education materials for students with print disabilities



---

# **Access and Use** ***or Beyond the Paywall***

---



# Viewing and Downloading

---

## Anyone Anywhere Can:

- Read public domain and open access works (via web)
- Search across the full text of the entire collection (via web)
- Receive a dataset of public domain content (under certain conditions)
- Mine text and data (via HTRC)

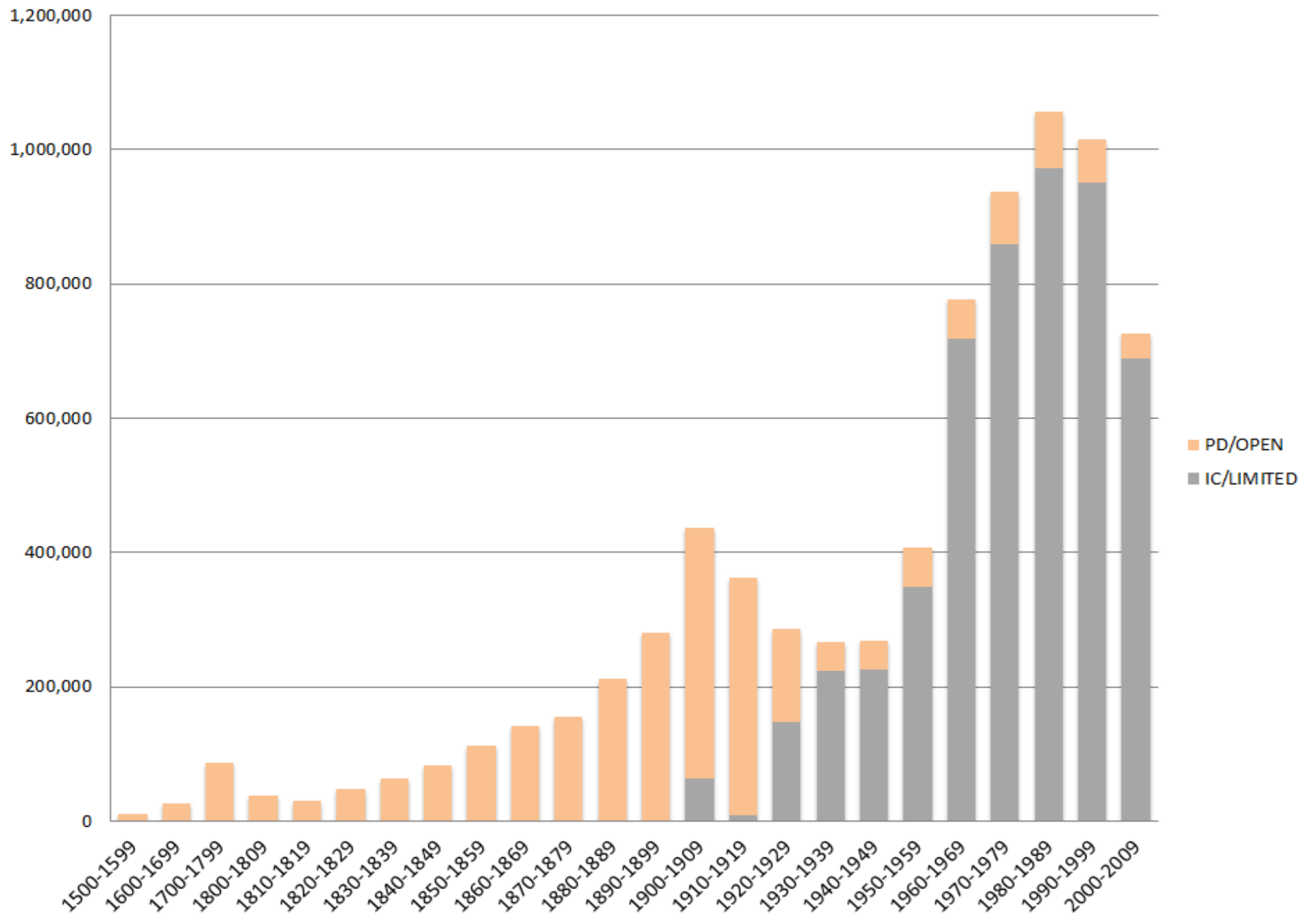
## Members Can:

- Download public domain and open access works
- Provide access to their users who are blind or with print disabilities
- Get access to replacement copies for lost and damaged print copies (under certain conditions)
- Mine text of copyrighted content (in pilot mode via HTRC)





# Distribution by Pub Date/Rights Status in HathiTrust, March 2017



# Providing Access to All Users

---

- Accessibility
  - HathiTrust Accessible Text Request Service
  - National Federation of the Blind (NFB) User Access to HathiTrust Copyrighted Collection
  - Marrakesh Treaty



# International Access and Use

---

- Membership for International Libraries
- International Access
- Full PDF Downloads



# Future Things

---

- Data Analysis and Collection Management Support
- Investigate Discovery to Delivery for Print Materials
- Policy on Metadata Sharing and Use
- Expanding Lawful Access to Works



# Mission and Vision

---

## Mission

The mission of HathiTrust is to contribute to research, scholarship, and the common good by collaboratively collecting, organizing, preserving, communicating, and sharing the record of human knowledge.

## Vision

HathiTrust will be a vital catalyst for emerging forms of research, teaching, and learning that engage the scholarly and cultural record.



# Questions & Thank You!

---

Heather Weltin

HathiTrust Shared Print Program Officer

[weltin@hathitrust.org](mailto:weltin@hathitrust.org)

