HATHITRUST

DIGITAL LIBRARY

HathiTrust and the University of Maryland

An Update

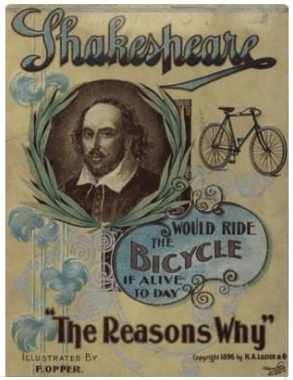
Mike Furlough

20 April 2016

University of Maryland

Sarah Werner @wynkenhimself — Follow

Would Shakespeare ride a bicycle? All answers in this 1896 @hathitrust pamphlet hdl.handle.net/2027/udel.3174… ht @AngelinaNZ



The Liz @lizwason — Follow

wow, @hathitrust provides us with this outlandishly lovely gem: a coloring book for dentists! bit.ly/1AjnHIo

The saga of Dr. Anton Pazu : a coloring book for dentists.
The saga of Dr. Anton Pazu : a coloring book for dentists.
babel.hathitrust.org

Jonathan Rees @jhrees — Follow

I can get access to all of HathiTrust from my hotel room. Good news for my book. Bad news for my sleep habits.

Donna Taylor @D_MT66 — Follow

I found one of my key documents digitized by @hathitrust from @nypl - that's saved me a lot of research time and train fare!

Gregg French @GreggFrench1 — Follow

Big thanks to @hathitrust for making my life just a bit easier. #history #academia #dissertation #writing

Jennifer Wondracek @JenWondracek — Follow

Love that HathiTrust is an integral part of UC's plan. #gpodlc15

American Libraries @amlibraries — Follow

All 754 vols of the NUC Pre-1956 Imprints are available online via HathiTrust, @ALCTS: on.fb.me/1NRDfKo

Kendra K. Levine @tranlib — Follow

To see how somethings never change, read John Cotton Dana's "Library Problems" hdl.handle.net/2027/njp.32101… Thank you @hathitrust.

Gillian Glaes @eurohistorian — Follow

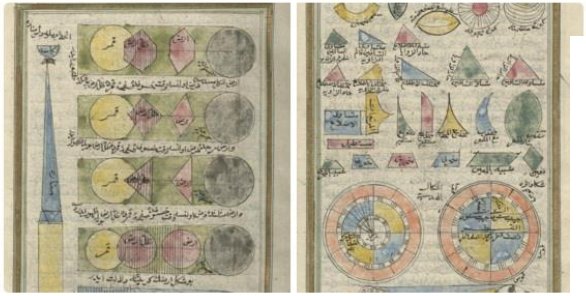
Hathitrust: an amazing digital library and resource collection: hathitrust.org #hstr3912 #hstr400 #hc12181

3:56 PM - 1 Oct 2015

mare nostrum @marenostrum2 — Follow

イブラヒム・ハック（1703-1780）知識の書
hdl.handle.net/2027/mdp.39015…

View translation



Klaus Riede @KlausRiede — Follow

Happy to see a fully digitised version of my dissertation online, thanks to google + Hathitrust hdl.handle.net/2027/uc1.b2714… Werbegesang G. rufus

Einfluss des werbegesangs von gomphocerus rufus l. auf...
Einfluss des werbegesangs von gomphocerus rufus l. auf das weib...
babel.hathitrust.org

Julia King @julialilinoe — Follow

This historian loves tech! Thanks @internetarchive and @hathitrust who saved my paper from death by early closing hours. #yesallmedievalists

Mission and Purpose

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

- *A trusted digital preservation service enabling the broadest possible access worldwide.*
- *An organization with over 100 research libraries partnering to develop its programs.*
- *A range of transformative programs enabled by working at a very large scale.*

HathiTrust remains **one of the best things that has happened to our library in the last 10 years**. Thank you.

One of our motivations is that **HathiTrust is a COLLECTIVE solution with access**.

...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 HathiTrust and doing more appropriate local things.

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.

We see **the advantages of working to scale even if our local infrastructure is adequate.**

The Name

- The meaning behind the name
  - Hathi (hah-tee)--Hindi for elephant
  - Big, strong
  - Never forgets, wise
  - Secure
  - Trustworthy

HE Law of the Jungle —which is by far the oldest law in the world—has arranged for almost every kind of accident that may befall the Jungle-People, till now its code is as perfect as time and custom can make it. You will remember.

Illustration of “Hathi” the elephant from 1895 edition of The Jungle Book found in HathiTrust.

Today's Discussion

- The organization
- What we've collected
- Access services
- Strategic initiatives

- Your questions answered

The Organization

HathiTrust Members

Allegheny College
American University of Beirut
Arizona State University
Auburn University
Baylor University
Boston College
Boston University
Brandeis University
Brown University
Bryn Mawr College
Carnegie Mellon University
Case Western Reserve
Colby College
Columbia University
Cornell University
Dartmouth College
Duke University
Emory University
Getty Research Institute
Georgetown University
Georgia Tech
Harvard University Library
Haverford College
Indiana University
Iowa State University
Johns Hopkins University
Kansas State University
Lafayette College
Library of Congress
Massachusetts Institute of Technology
McGill University`
Michigan State University
Montana State University
Mount Holyoke College
New Mexico State University
New York Public Library
New York University
North Carolina Central University
North Carolina State University
Northeastern University
Northwestern University
Oklahoma State University
The Ohio State University
The Pennsylvania State University
Princeton University
Purdue University
Rutgers University
Smith College
Stanford University
State University System of Florida
Swarthmore College
Syracuse University
Temple University
Texas A&M University
Texas Christian University
Texas Tech University
Tufts University
Universidad Complutense de Madrid
University of Alabama
University of Alberta
University of Arizona
University of British Columbia
University of Calgary
University of California
- Berkeley
- Davis
- Irvine
- Los Angeles
- Merced
- Riverside
- San Diego
- San Francisco
- Santa Barbara
- Santa Cruz
- California Digital Library

The University of Chicago
University of Connecticut
University of Delaware
University of Houston
University of Illinois
University of Illinois at Chicago
The University of Iowa
University of Kansas
University of Maine
University of Maryland
University of Massachusetts, Amherst
University of Miami
University of Michigan
University of Minnesota
University of Missouri
University of Nebraska-Lincoln
University of Nevada-Las Vegas
University of New Mexico
The University of North Carolina at Chapel Hill
University of Notre Dame
University of Oklahoma
University of Pennsylvania
University of Pittsburgh
University of Queensland
University of Rochester
University of Tennessee, Knoxville
University of Texas
University of Utah
University of Vermont
University of Virginia
University of Washington
University of Wisconsin-Madison
University of Wyoming
Utah State University
Vanderbilt University
Virginia Tech
Wake Forest University
Washington University
Washington State University
Wesleyan University
Williams College
Yale University Library

Membership Activities

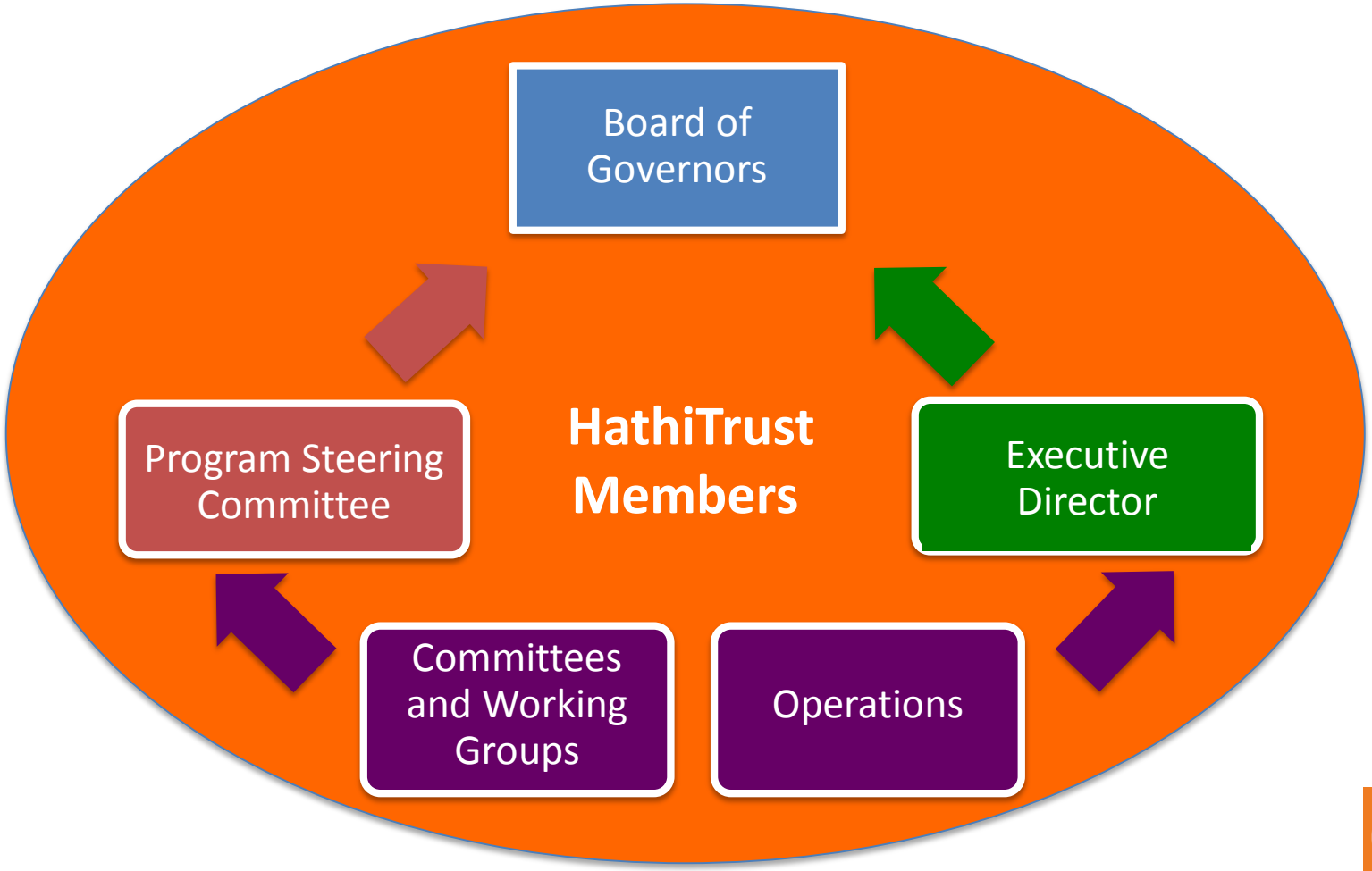
- Annual member meeting
  - Nov 10, 2016, Big Ten Center, Chicago IL
- Vote on budget, Board of Governors, etc.
- Submit print holdings
  - Required annually
- Take advantage of HT services locally
  - Requires Shibboleth implementation
- Deposit your content
- Participate in working groups
- Participate in current initiatives

Cooperative Work

We draw upon distributed expertise

	Michigan	Indiana	Illinois	California
Administration	✓			
Preservation & Access Repository	✓	✓		
Research Center		✓	✓	
Metadata Management (Zephir)				✓

Governance


Board of Governors
Executive Director
Operations
Committees and Working Groups
Program Steering Committee
HathiTrust Members

Currently Active Committees

- Board of Governors
- Program Steering Committee
- Collections Committee
- Metadata Policy, Strategy Use and Advisory Group
- Zephir Advisory Group
- Rights & Access Working Group
- Ad Hoc Quality Working Group
- User Support Working Group*

- Forthcoming
  - Print Monographs Archive Advisory Group
  - Federal Documents Advisory Group

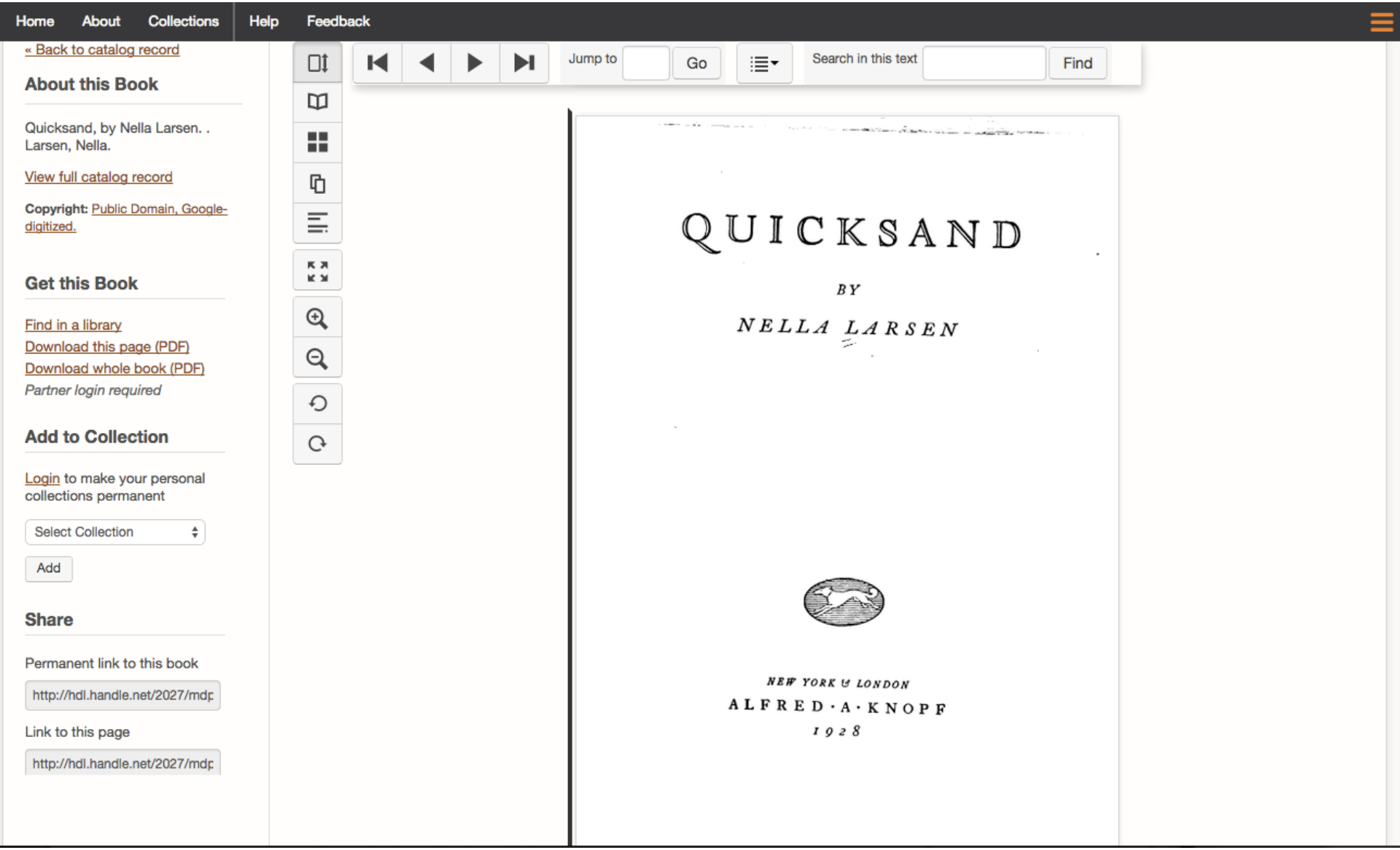
The Collection

HathiTrust in April 2016

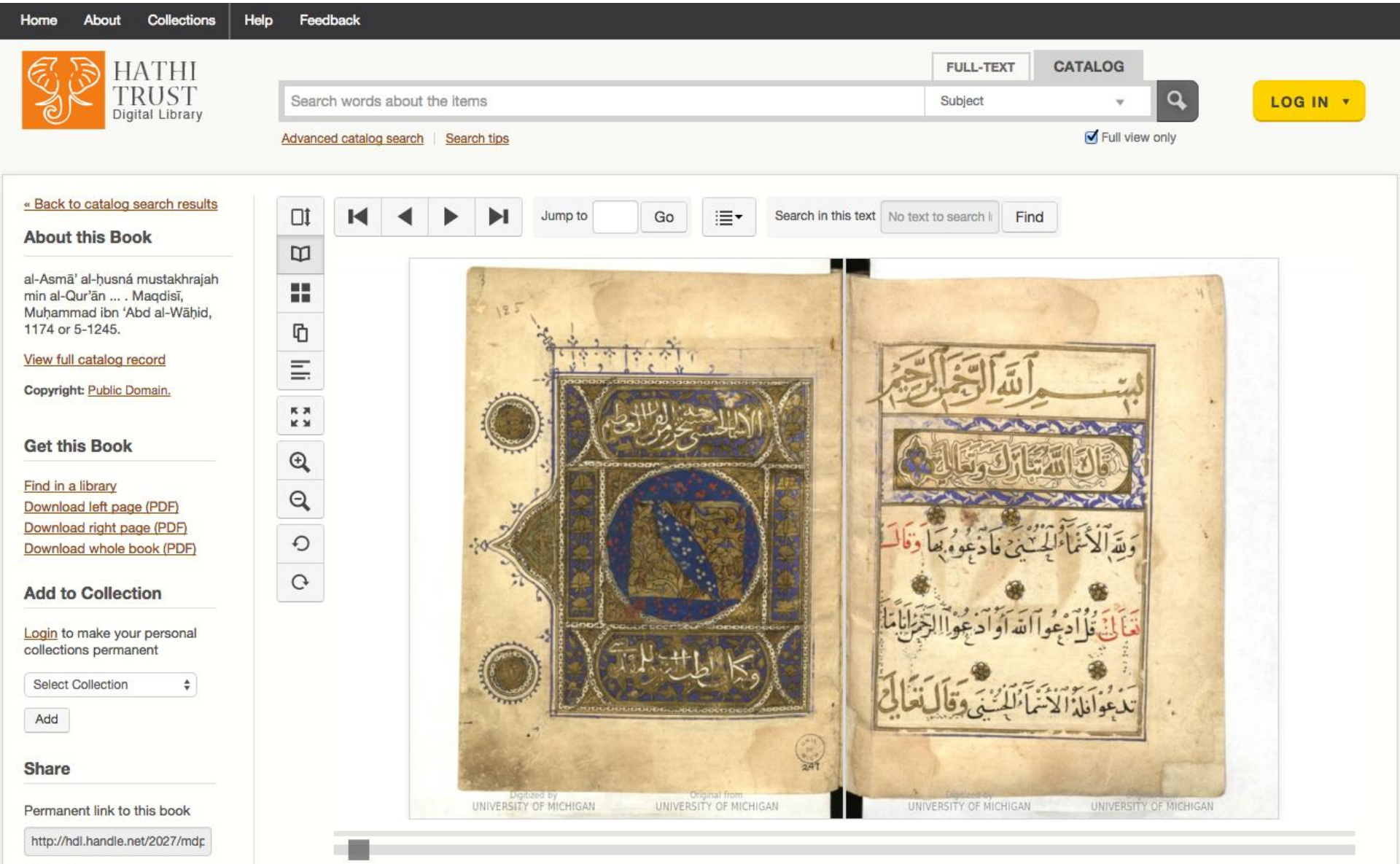
- 14 million total items
  - 7 million book titles
  - 371,000 serial titles
  - 726,000 US federal government documents
  - 5.5 million items open (public domain & CC-licenses)

The collection includes (mostly) published materials in bound form, **digitized from library collections.**

Published Items: Google Digitized



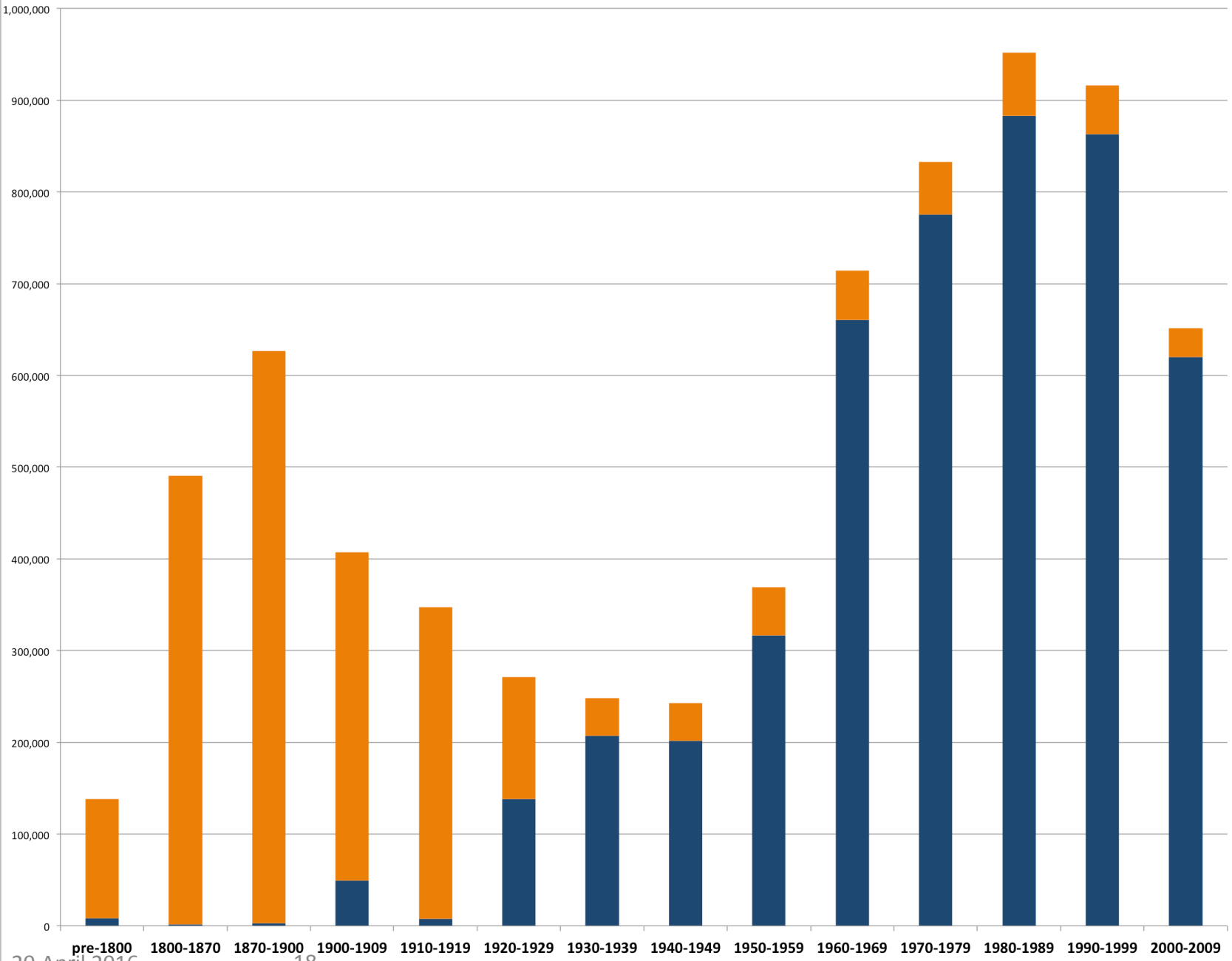
Special Collections: Islamic Manuscripts



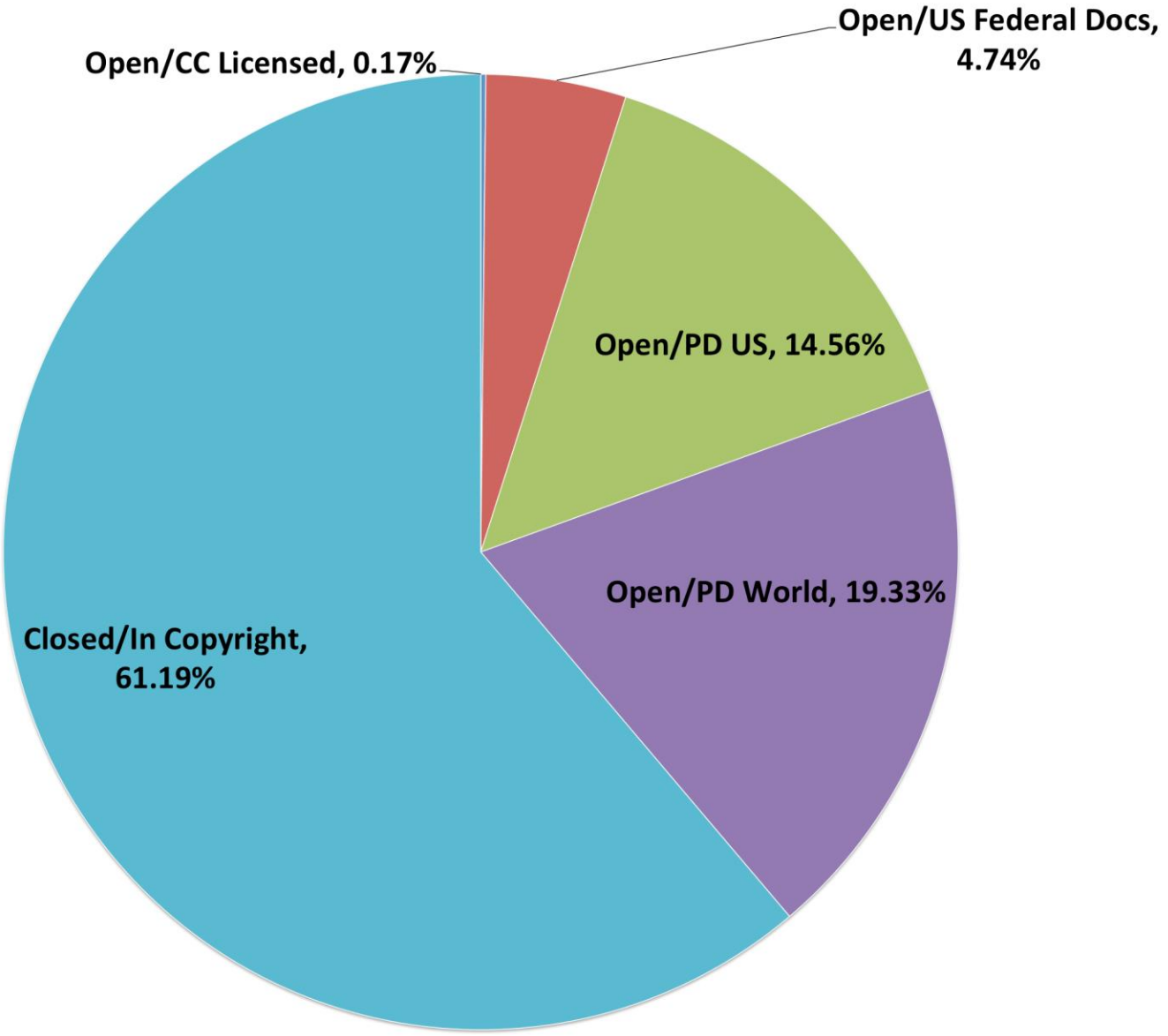
Collection Contributors

Boston College	New York Public Library	University of Connecticut
Brooklyn Museum	North Carolina State University	University of Delaware
Clark Art Institute Library	Northwestern University	Univ of Illinois at Urbana-Champaign
Columbia University	Pennsylvania State University	University of Iowa
Cornell University	Princeton University	University of Maryland, College Park
Duke University	Purdue University	University of Massachusetts Amherst
Emory University	State University System of Florida	University of Michigan
Harvard University	Technical Report Archive & Image Library	University of Minnesota
Indiana University	Texas A&M University	University of Missouri - Columbia
Keio University	The Getty Research Institute	Univ of North Carolina at Chapel Hill
Knowledge Unlatched	The Ohio State University	University of Virginia
Library of Congress	Tufts University	University of Washington
McGill University	Universidad Complutense de Madrid	University of Wisconsin - Madison
Michigan State University	University of Alberta	Utah State University Press
Millennium Project	University of California	Wake Forest University
Minnesota Digital Library	University of Chicago	Yale University

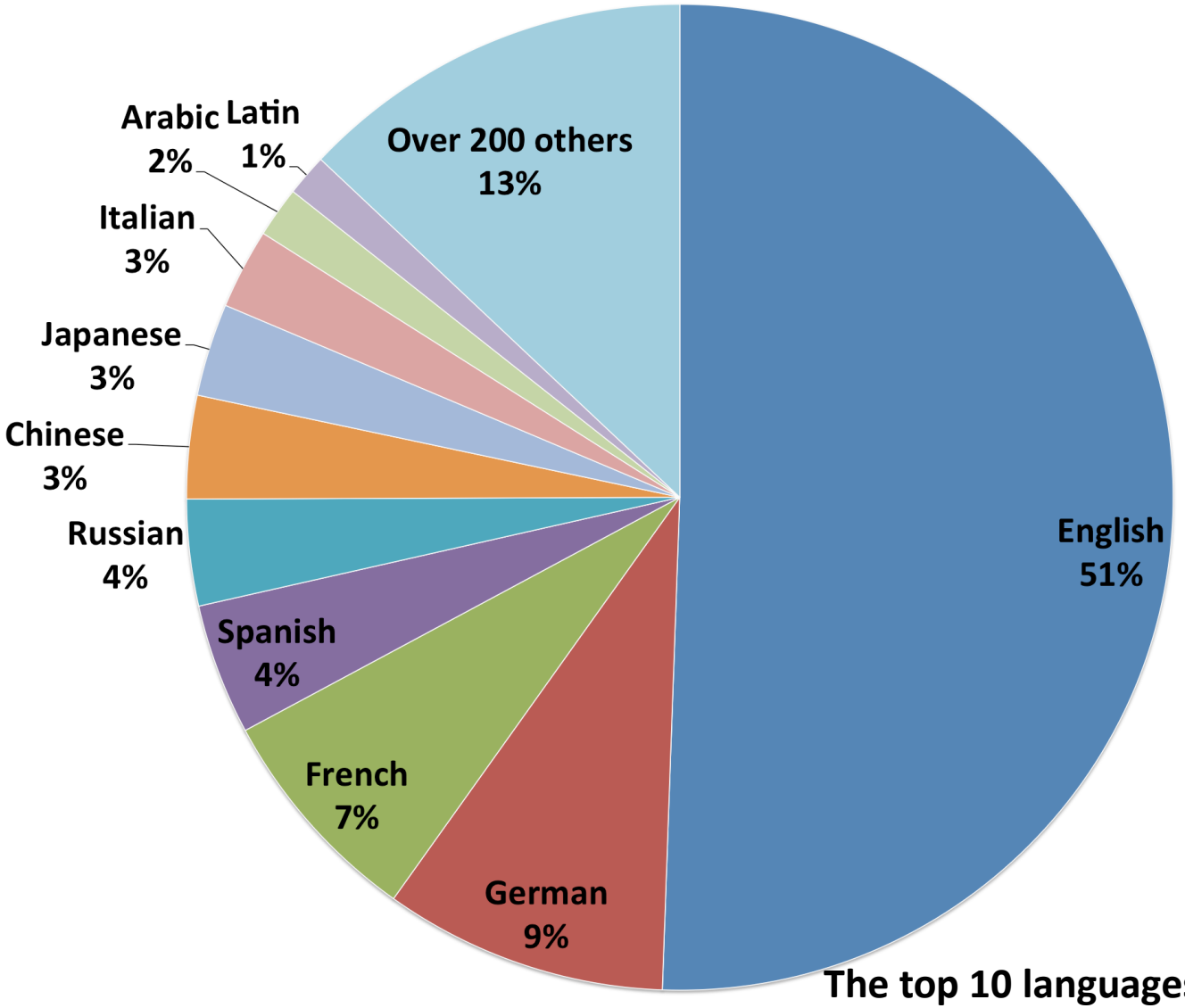
HathiTrust Titles by Date and Viewing Status



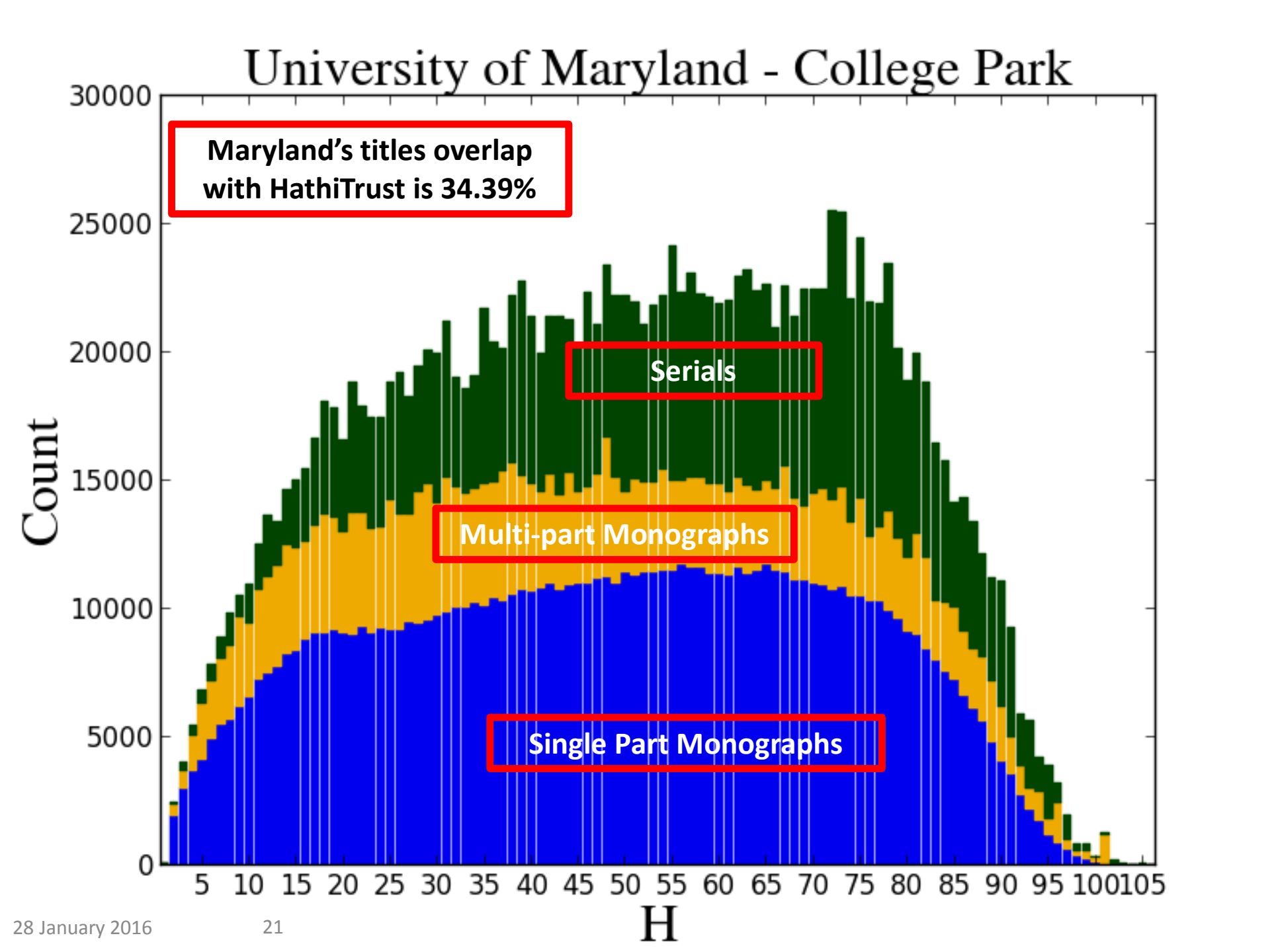
HathiTrust Titles by Copyright/View Status



Titles by Language in HathiTrust -- all items



The top 10 languages make up ~87% of all content


University of Maryland - College Park
Maryland's titles overlap with HathiTrust is 34.39%
Serials
Multi-part Monographs
Single Part Monographs
Count
H
30000
25000
20000
15000
10000
5000
0
5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100105

Top Ten Titles 2015

1. *Roster of the Confederate soldiers of Georgia, 1861-1865 (v.1-5)*
2. *Solid mensuration*, by Willis F. Kern and James R. Bland.
3. *Quicksand*, by Nella Larsen.
4. *Godey's magazine,* v.40-41 1850
5. *A Short Guide to New Zealand*, [Special Services Division, US Army]
6. *The human figure*, by John H. Vanderpoel.
7. *Pennsylvania German Pioneers: a publication of the original lists of arrivals in the port of Philadelphia from 1727 to 1808*
8. *Quintus Curtius [History of Alexander]* trans. By John C. Rolfe
9. *Modern California houses; case study houses, 1945-1962,* by Esther McCoy.
10. *Consumption of the lungs and kindred diseases, treated and cured by kerosene,* by Charles Oscar King

Collections:
Access Services

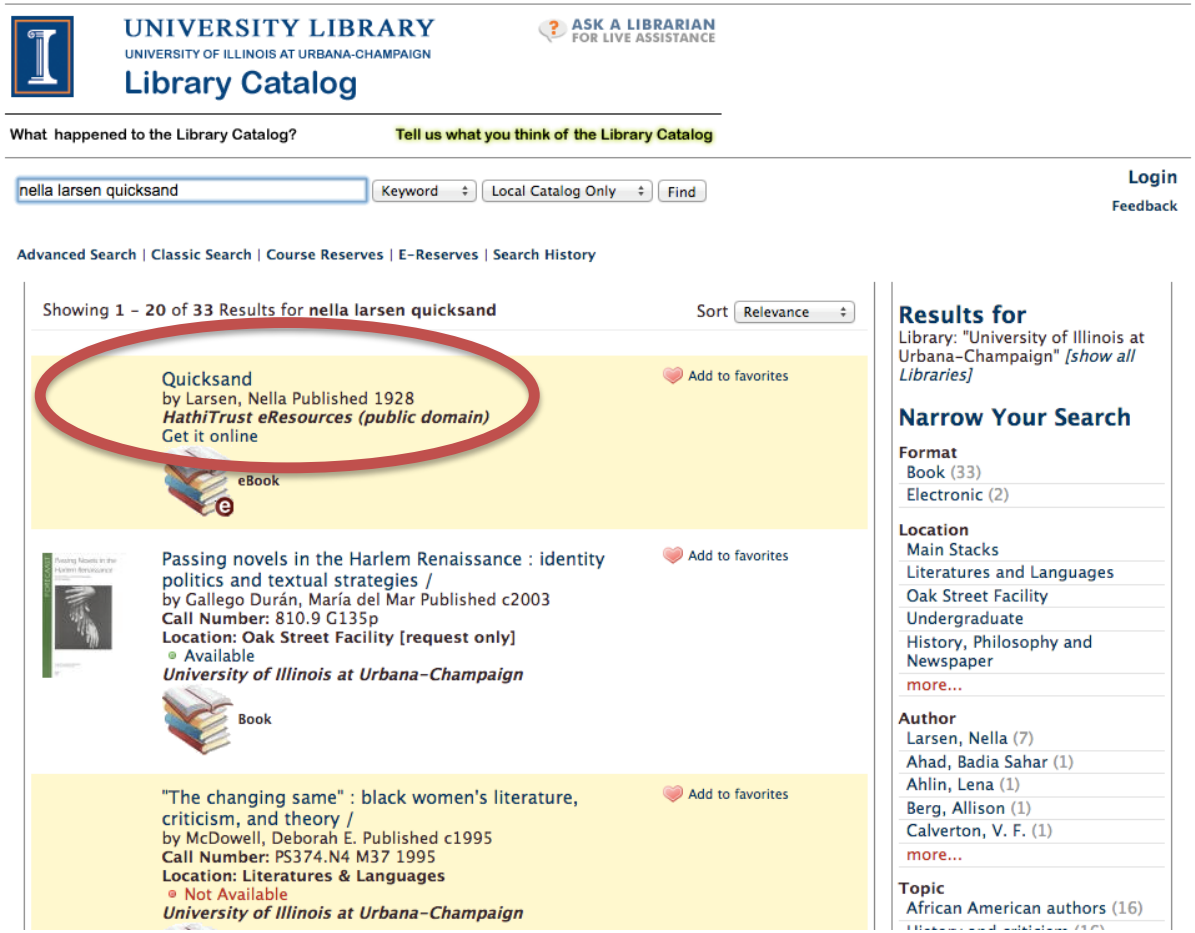
Preservation with Access

- Preservation
  - TRAC-certified
  - Replication node for Digital Preservation Network (DPN)
- Discovery
  - Bibliographic and full-text search of all materials (all)
  - Bibliographic data API enables local catalog integration (all)
- Access and Use
  - Varies

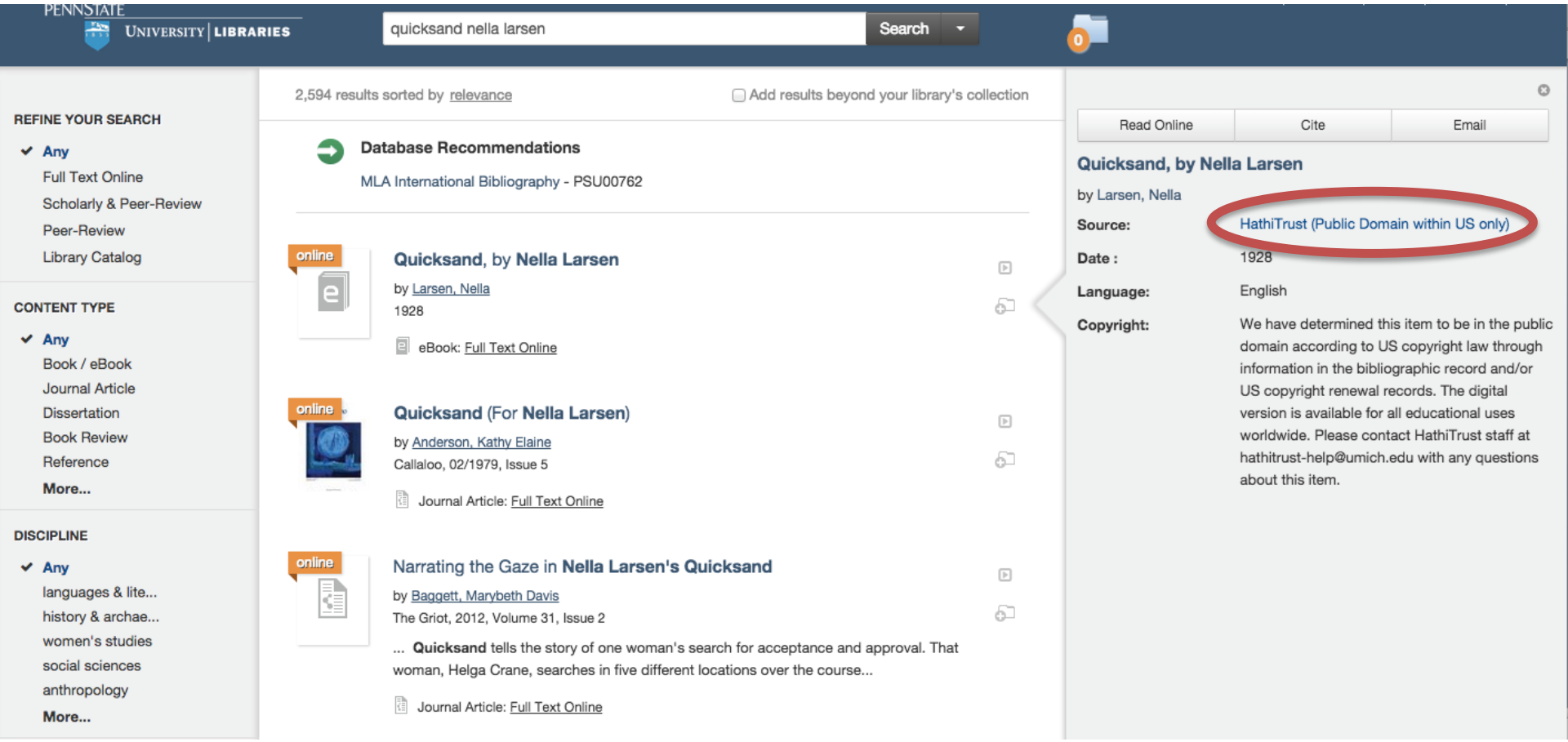
Modes of Access & Use

- Full text search (all users)
  - Text and data mining (all users)
    - Services require additional registration
- Public domain and open access works (all users)
  - Full-item download (members only)
- Collection building and APIs (all users)
- Lawful uses of in-copyright works (members only)
  - Access for users with print disabilities
  - Preservation replacement

Finding via Local Catalog



Finding via Discovery Service



Top Referring Sites

Website	Sessions	% New Sessions
google.com	4,723,224	68.62%
onlinebooks.library.upenn.edu	911,699	19.88%
worldcat.org	262,240	5.72%
dp.la	60,117	1.31%
mirlyn.lib.umich.edu	54,180	1.18%
clio.columbia.edu	40,821	0.89%
en.wikipedia.org	38,806	0.85%
psu.summon.serialssolutions.com	31,452	0.69%
vufind.carli.illinois.edu	31,078	0.68%
oskicat.berkeley.edu	30,510	0.27%
lccn.loc.gov	29,289	0.27%

Period: 12/1/2014 to 11/30/2015. Data taken from Google Analytics.

Bringing Buried Treasure to Light

article-level discovery metadata for HathiTrust

Michelle Paolillo *presenting for John Cline, Steven Folsom, Jaron Porciello*

Cornell University

HathiTrust Annual Meeting
December 9, 2015

Enhanced Access to HathiTrust for Patrons with Disabilities: Experiences of the State University Libraries of Florida

Ben Walker
Assistant Dean, Digital Services & Shared Collections
George A. Smathers Libraries
University of Florida

Current Initiatives

Current Major Cooperative Initiatives

- Distributed manual copyright reviews.
- Establishing a distributed shared print monograph archive.
- Expanding services of the HathiTrust Research Center.
- Expanding and enhancing access to US Federal Government Documents

Collective Action: Copyright Review

- Copyright Review Management System
  - Systematic manual review of copyright registrations to determine status of portions of the HathiTrust Collection, Supported generously by IMLS

Project	**Reviewed**	**Out of Copyright**
CRMS US Pub in US 1923-1963	331,899	177,398 (53%)
CRMS World Pub in UK (1875-1944), Canada and Australia (1894-1964)	272,300	144,733 (53%)

Federal Documents Initiative

- Ballot Initiative passed at the 2011 HT Constitutional Convention
- Focus:
  - expanded coverage & enhanced access to U.S. federal documents.
- Near term activities:
  - Developing a registry of US Federal Government Documents
    - https://www.hathitrust.org/usdocs_registry
  - Digitize! Focus first on known and cataloged materials
    - Gap analysis driven
    - Prioritize print, post-1976 materials
  - Publicize the efforts
    - Within the library community and the general public

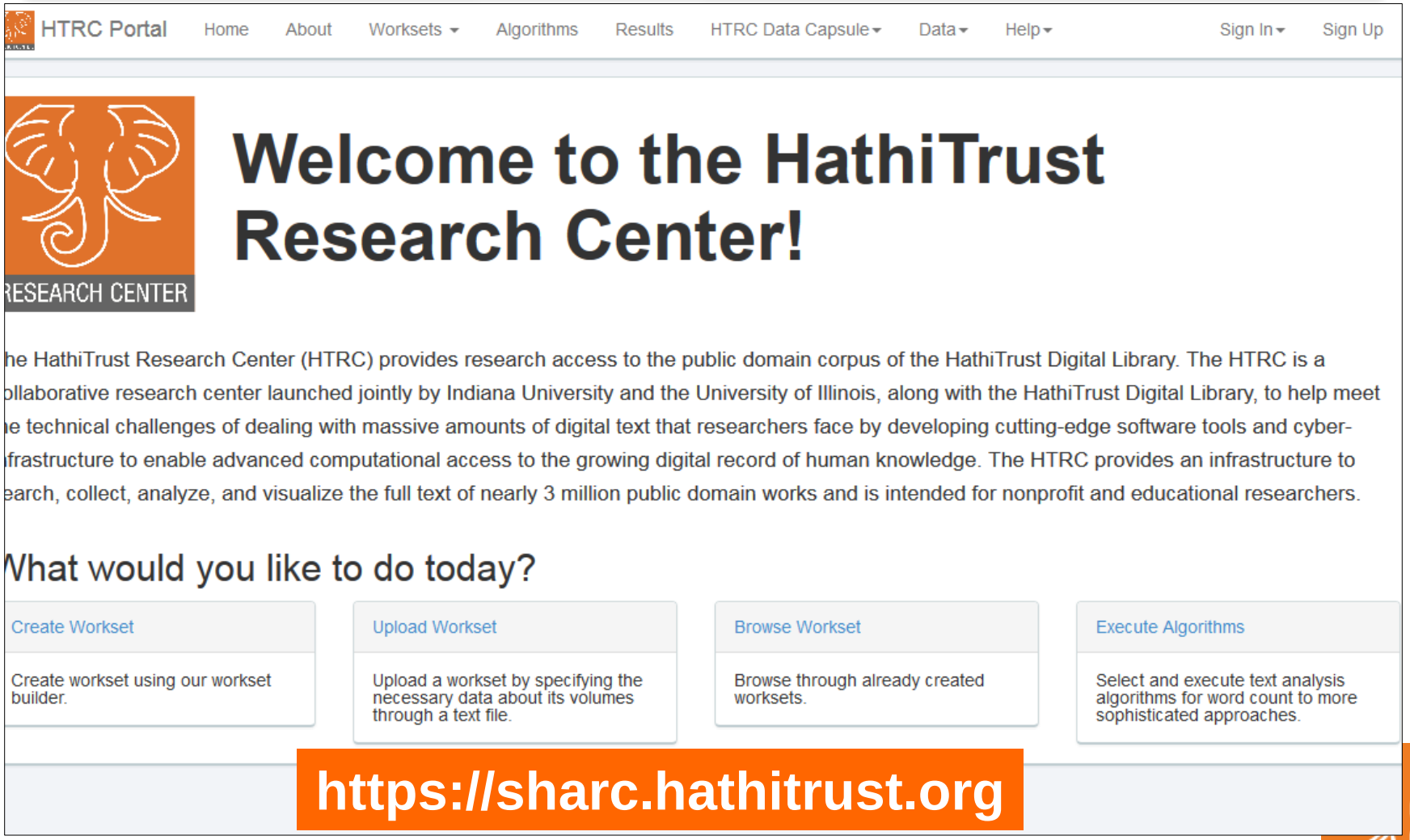
Shared Print Monographs Archive

- Ballot Initiative passed at the 2011 HT Constitutional Convention
- Focus
  - Ensure preservation of print and digital collections
  - Catalyze national/continental collective management of collections
- Proposed Features:
  - Secure retention commitments for all monograph titles in HathiTrust. (50% within two years or less.)
  - Ensure that the retained volumes are accessible (will circulate).
  - Disclose commitments and discover content and support local collection management.
- Original proposal, task force charge and preliminary recommendations:
  - https://www.hathitrust.org/print_monograph_archiving

“Non-consumptive” Research: The HathiTrust Research Center

- HathiTrust Research Center
  - Persistent and sustainable structure to enable original and cutting edge *non-consumptive* research
  - Developed collaboratively by Indiana University and University of Illinois.
    - Additional funding from HathiTrust and foundations
  - Analytics Portal
    - https://sharc.hathitrust.org/
  - Advanced Collaborative Support Programs
    - https://www.hathitrust.org/htrc_sp16_acs-rfp
  - Dataset distribution:
    - https://www.hathitrust.org/datasets

HTRC Portal

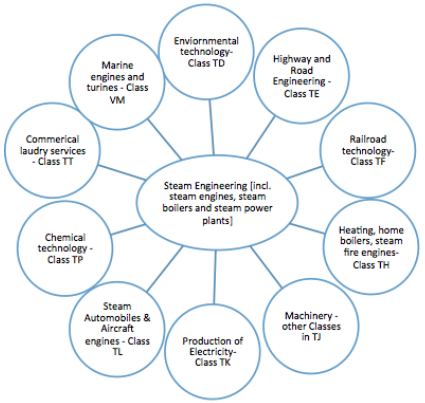

HTRC Portal
Home
About
Worksets
Algorithms
Results
HTRC Data Capsule
Data
Help
Sign In
Sign Up
RESEARCH CENTER
Welcome to the HathiTrust Research Center!
he HathiTrust Research Center (HTRC) provides research access to the public domain corpus of the HathiTrust Digital Library. The HTRC is a
ollaborative research center launched jointly by Indiana University and the University of Illinois, along with the HathiTrust Digital Library, to help meet
e technical challenges of dealing with massive amounts of digital text that researchers face by developing cutting-edge software tools and cyber-
frastructure to enable advanced computational access to the growing digital record of human knowledge. The HTRC provides an infrastructure to
earch, collect, analyze, and visualize the full text of nearly 3 million public domain works and is intended for nonprofit and educational researchers.
What would you like to do today?
Create Workset
Create workset using our workset builder.
Upload Workset
Upload a workset by specifying the necessary data about its volumes through a text file.
Browse Workset
Browse through already created worksets.
Execute Algorithms
Select and execute text analysis algorithms for word count to more sophisticated approaches.
https://sharc.hathitrust.org

Example Advanced Collaborative Support Projects

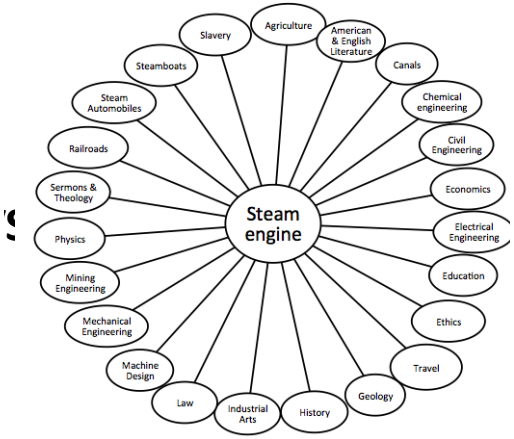
Tracking Technology Diffusion Through Time in the HathiTrust Corpus

Michelle Alexopoulos, University of Toronto

Dr. Alexopoulos, an economist, is using the vast historical record contained in the HathiTrust to study the diffusion of various technologies over time. By tracking word usage trends of 1,214 technology-related terms identified by Alexopoulos, such as the steam engine, her research based on HathiTrust book content has the potential to overturn accepted theories about the economic and societal impacts of a technology.



Linkages to "Steam Engines" implied by the Library of Congress Classification



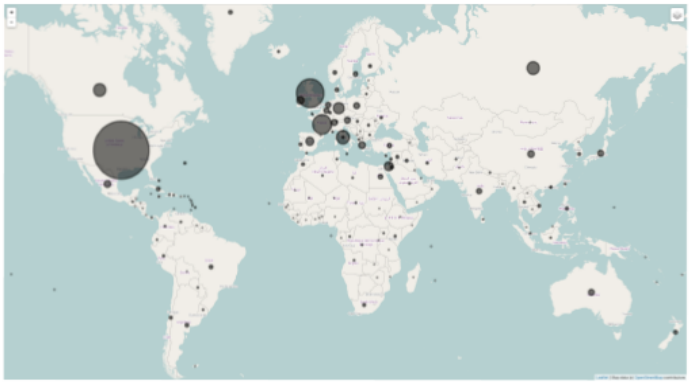
From HT text: Selected subject terms linked to "Steam engine" n-gram by 1910

- 1,012,633 volumes analyzed.
- Over 22 hours of processing using a 32-node cluster on Indiana University's high-performance supercomputer, Big Red II.
- Each node had 32 cores and 64 GB of RAM.

Literary Geography at Scale

Matthew Wilkens, University of Notre Dame

With the help of natural language processing, Dr. Wilkens will extract and geocode place names in nearly eleven million volumes from the HathiTrust, including twentieth and twenty-first century texts. This project to geolocate world literature, supported by a 2014-2015 American Council for Learned Societies Digital Innovation Fellowship, is one of the largest humanities text-mining projects to date.



- Pilot analysis completed and named entity extraction run on 10,000 randomly sampled volumes.
- Pilot took 55 minutes of processing time on Indiana University's high-performance computer, Big Red II, using an 8-node cluster.
- Next step: Process the entire corpus!

Wrapping Up

2016 Agenda

- Launch of Shared Print Monographs Program.
- Active collection development in Federal Documents effort.
- Improving engagement across the membership.
  - Input into development agenda (ballot initiative)
- Transitioning copyright review to ongoing activity.
- Improvements to print disability services.

Assumptions

- Our actions must align with the mission, goals, and purpose across our partnership.
- A few additional assumptions
  - Internationally we should pursue complementarity and cooperation, not competition and duplication.
  - Scale will continue to drive our work
  - Potential partners are not just other libraries and library organizations, but also readers, authors, publishers.

How to find out more

- Contact us: feedback@issues.hathitrust.org

- About: http://www.hathitrust.org/about
- Resources: http://www.hathitrust.org/resources
- Twitter: http://twitter.com/hathitrust
- Facebook: http://www.facebook.com/hathitrust
- Monthly newsletter:
  - http:www.hathitrust.org/updates
- Blogs: http://www.hathitrust.org/blogs
  - Large-scale Search
  - Perspectives from HathiTrust

Thank You!

Your Questions