

Your Library, Now Online! Putting HathiTrust in the Context of Traditional (and New) Library Services

MCLS Webinar February 6, 2012 Jeremy York, Project Librarian, HathiTrust



Unless otherwise noted, these slides and their contents are licensed under a <u>Creative Commons Attribution Unported License</u>.

Outline

- The Big Idea
 - Mission and Goals
- · What we're doing to get there
 - Repository and Content
 - Making content available
 - Organizational structure
- How HathiTrust can change the way we work



The Big Idea



Partnership

North Carolina State

Arizona State University Baylor University Boston College Boston University Brandeis University California Digital Library Carnegie Mellon University Columbia University Cornell University Dartmouth College Duke University Emory University Florida State University Getty Research Institute Harvard University Library Indiana University
Iowa State University
Johns Hopkins University Kansas State University Kansas State University
Lafayette College
Library of Congress
Massachusetts Institute of
Technology
McGill University
Michigan State University
New York Public Library
New York University
North Carolina Central
Liniversity

University

University
Northwestern University
The Ohio State University
The Pennsylvania State University Princeton University
Purdue University
Stanford University
Syracuse University Syracuse University
Texas A&M University
Universidad Complutense
de Madrid
University of Arizona
University of Calgary
University of California
Berkeley
Davis Davis Irvine Los Angeles Merced Riverside San Diego San Francisco Santa Barbara Santa Cruz The University of Chicago University of Connecticut University of Delaware University of Florida

University of Illinois University of Illinois at University of Illinois at Chicago
The University of Iowa University of Kansas University of Maryland University of Michigan University of Michigan University of Minnesota University of Missouri University of Nebraska-Lincoln Lincoln
The University of North
Carolina at Chapel Hill Carolina at Chapel H
University of Notre Dame
University of Pennsylvania
University of Pittsburgh
University of Utah
University of Vermont
University of Virginia
University of Washington
University of Wisconsin University of Wisconsin-Madison Utah State University Vanderbilt University Vanderbilt Orliversity
Virginia Tech
Wake Forest University
Washington University
Yale University Library

Digital Repository

- Launched 2008
- Initial focus on digitized book and journal content
 - 10.6 million total volumes
 - 5.58 million book titles
 - 276,000 serial titles
 - 3.2 million public domain (~31%)



The Name

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

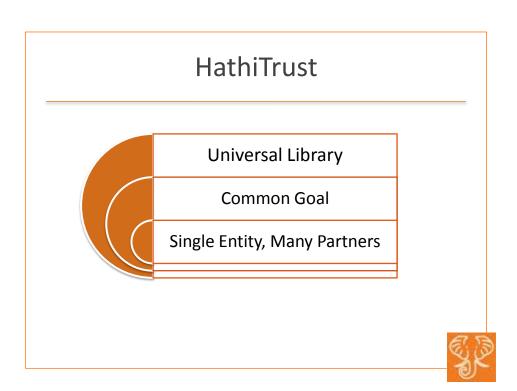




Mission

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





Collections and Collaboration

- Comprehensive collection
 - Preservation...with Access
- Shared strategies
 - Copyright
 - Collection management, development
 - Preservation
 - Discovery / Use
 - Bibliographic Indeterminacy
 - Efficient user services
- Public Good



What we are doing to get there



Cost-effective long-term preservation and access for digitized content

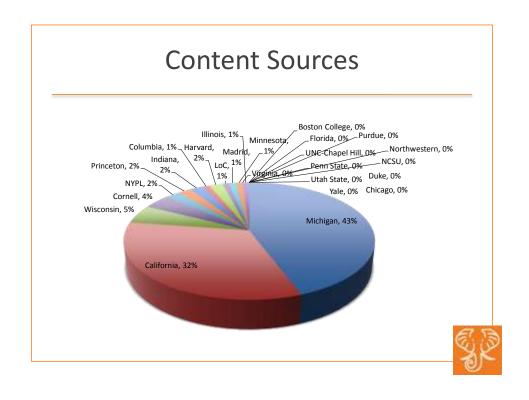


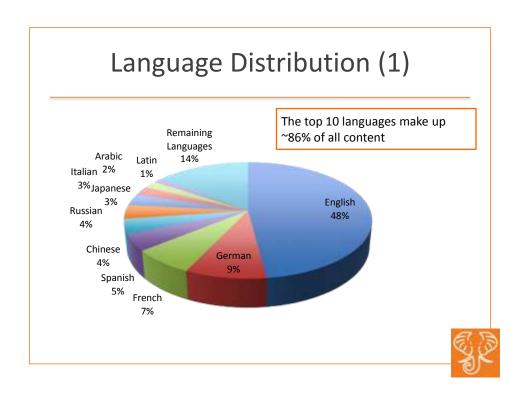
- Facilitate decision-making about digitization and print collection management
- Facilitate activities such as discovery, copyright review, use of materials

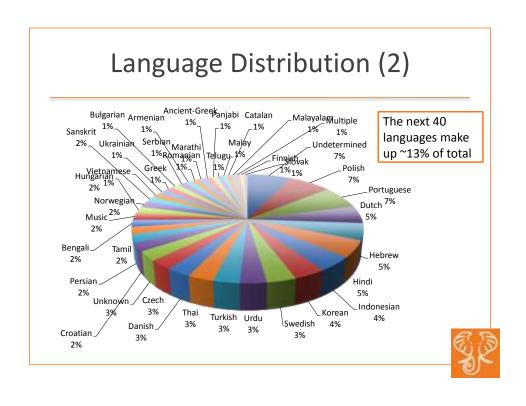


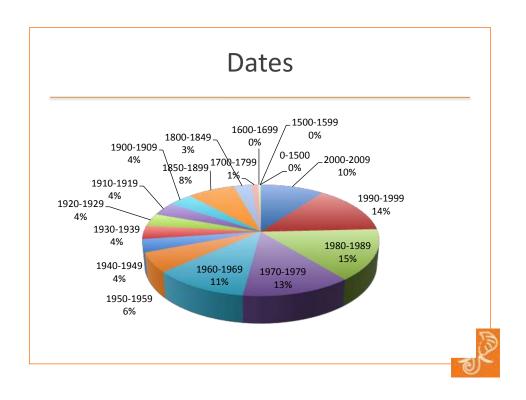
Repository and Content

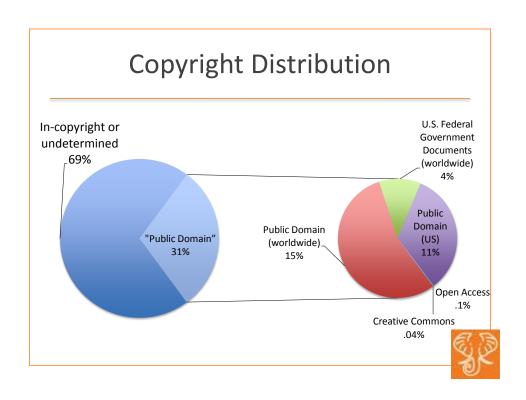


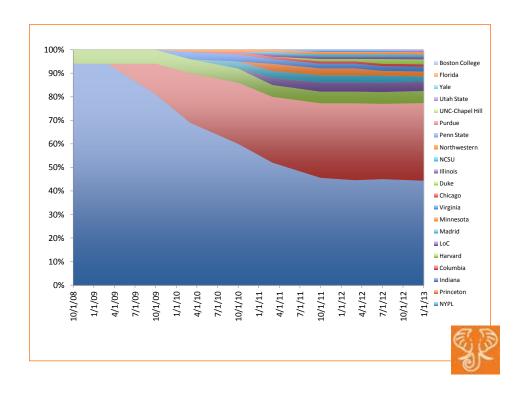


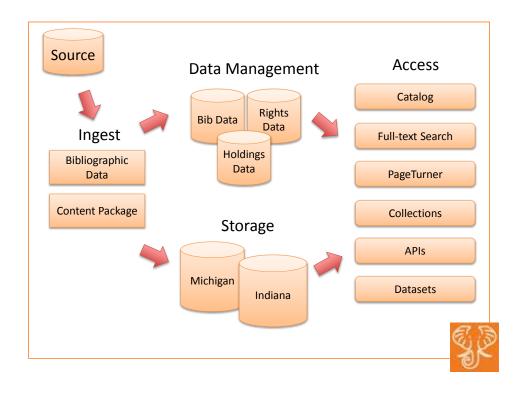


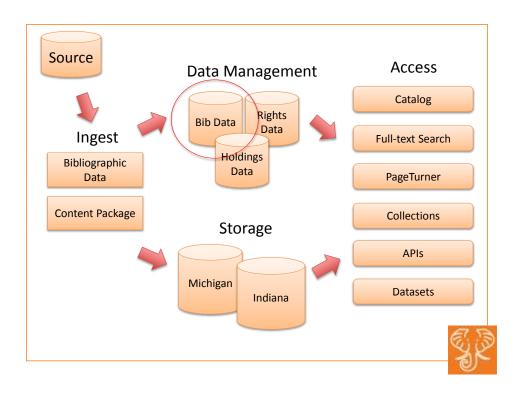


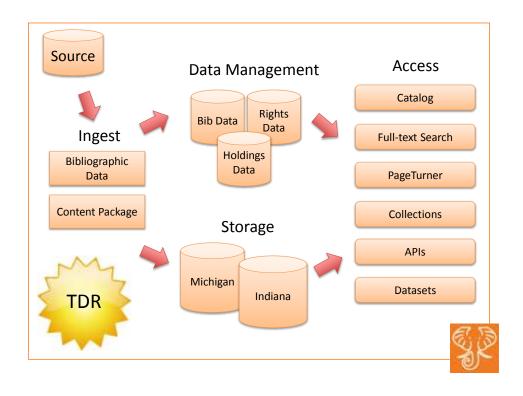


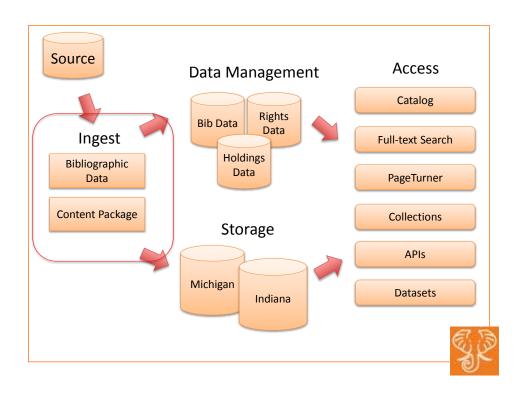


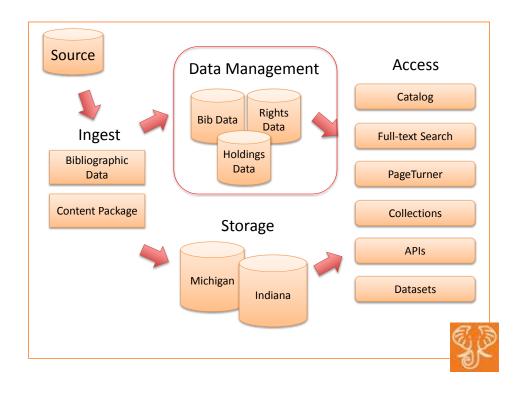


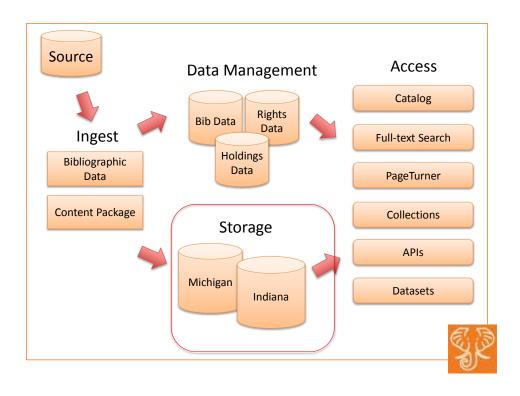


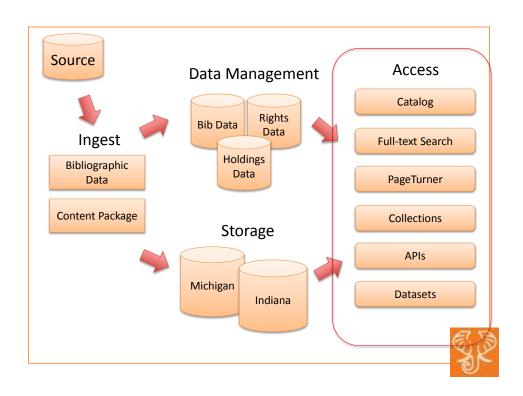






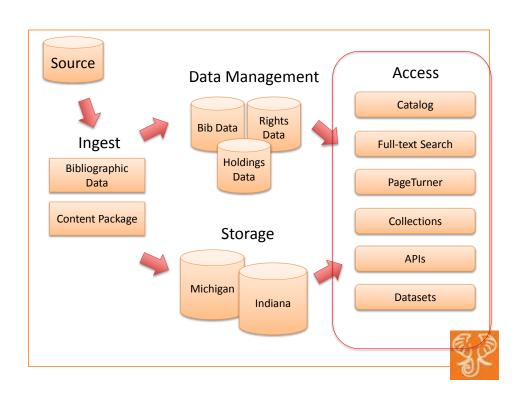






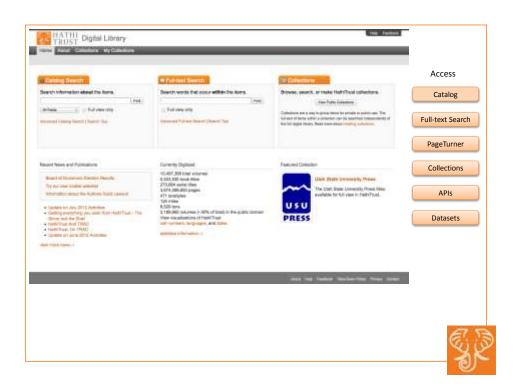
We engage in preservation for purposes of access

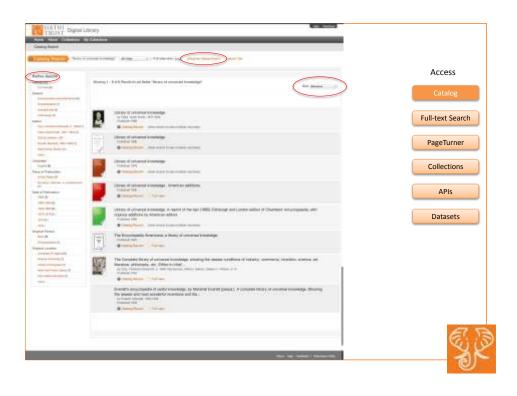


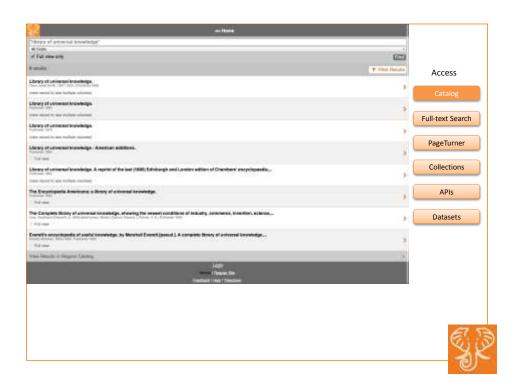


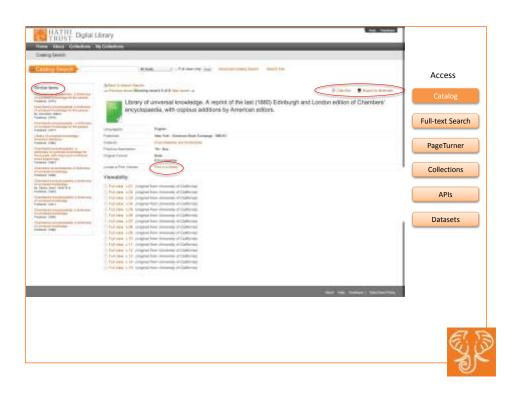
Making Content Available

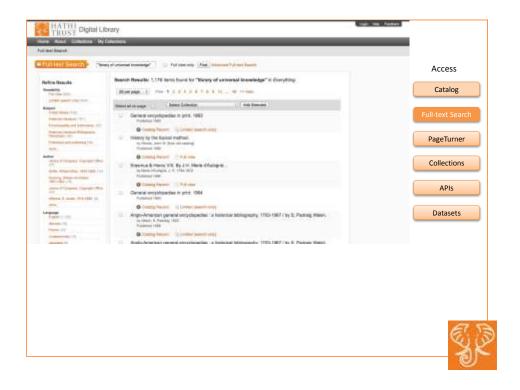


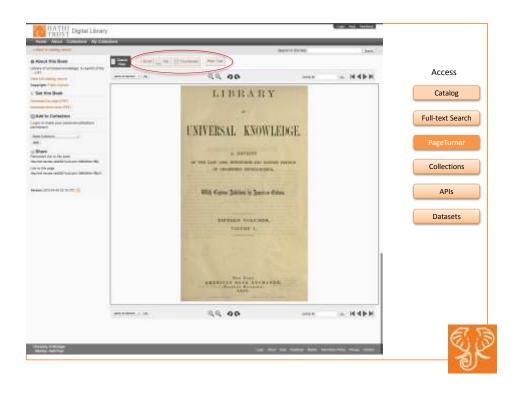


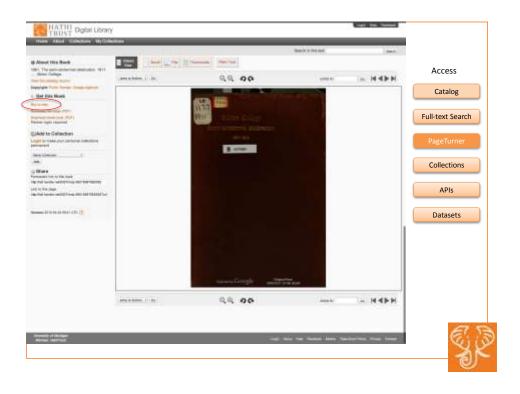


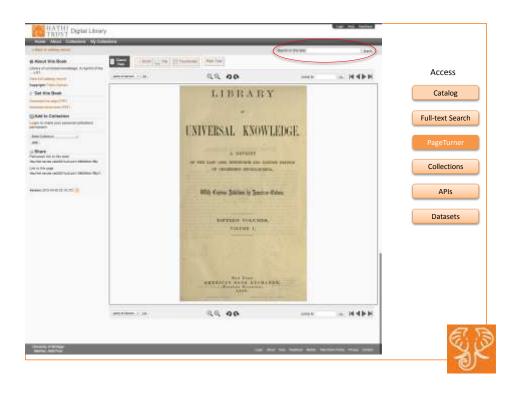


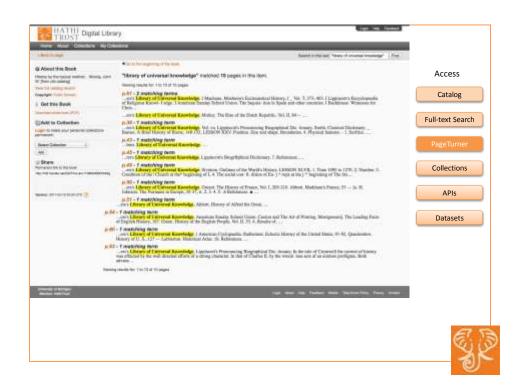


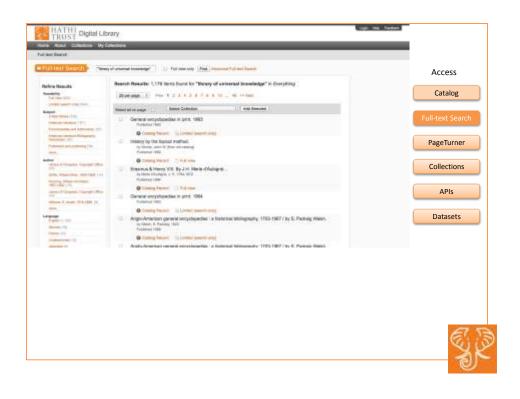


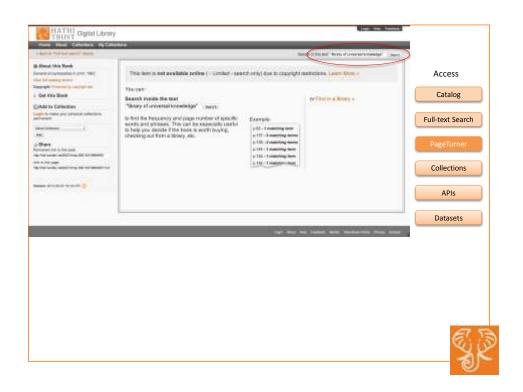


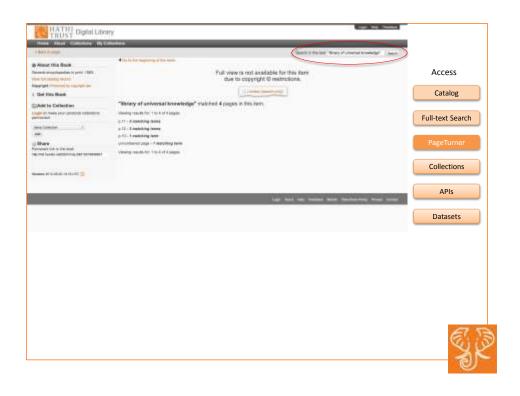


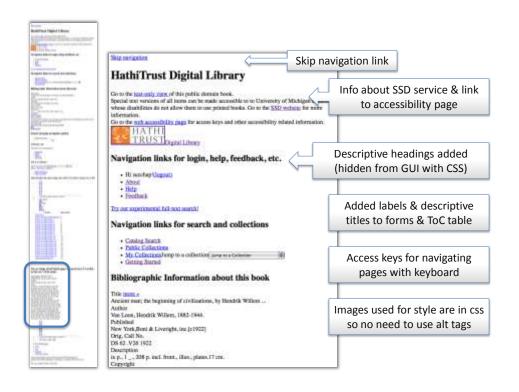


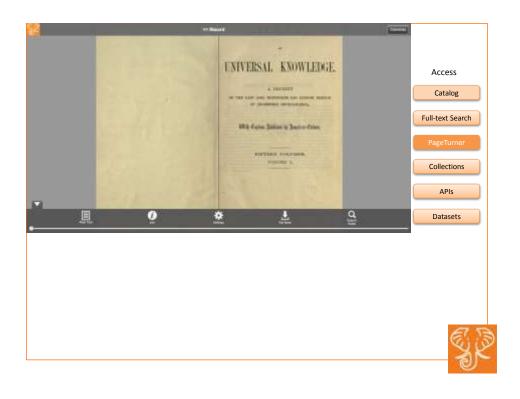


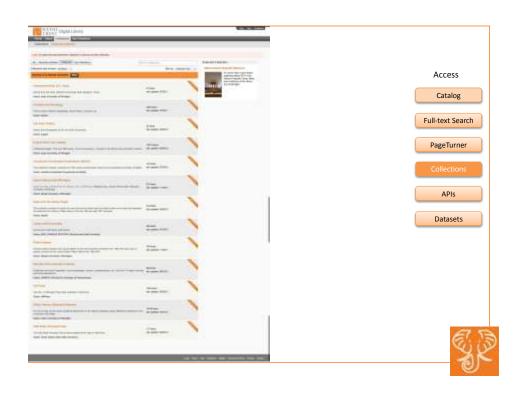


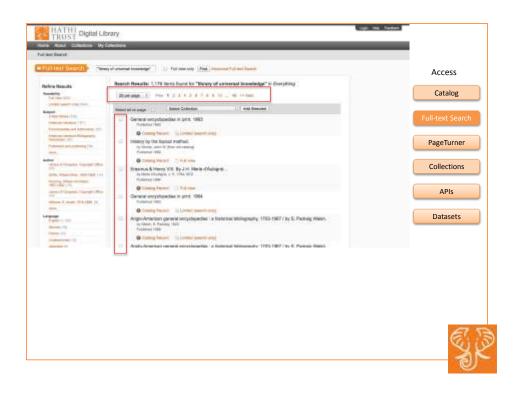


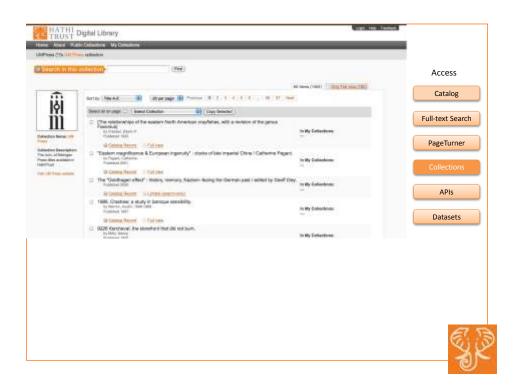








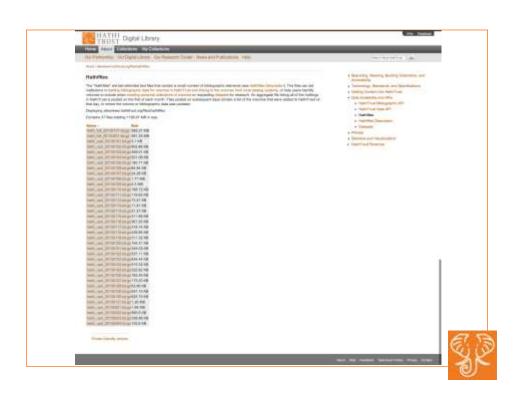




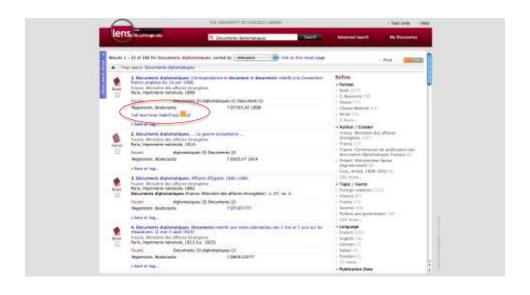
APIs

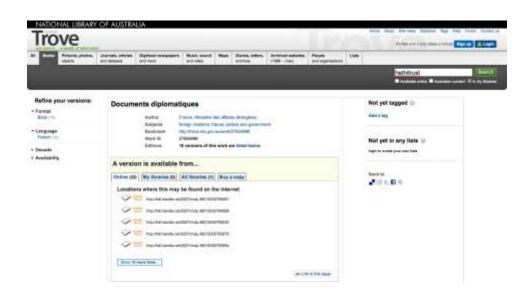
- Data API
 - Volume and rights information
 - Page images
 - OCR
- Bibliographic API
 - Volume and rights information
 - MARC records
- OAI
- "Hathifiles"











Datasets

- Google-digitized
 - ~2.8 million texts
 - Requires proposal to HathiTrust
 - Agreement with Google
 - Statement on use/management
- Non-Google-digitized
 - ~370,000 texts
 - Freely available
 - Statement on management



Research Center

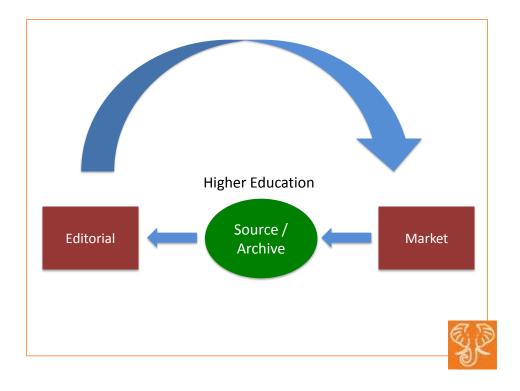
- Environment to perform research on HathiTrust corpus
- http://www.hathitrust.org/htrc



mPach

- http://lib.umich.edu/mpach
- Package of tools to enable publication of open access, born-digital journal content, directly into HathiTrust
 - Including accompanying data and media files
- Allows integration with popular journal publishing tools such as Open Journal Systems (OJS)





Access Determinations

- Automated
- Manual



Automatic Rights Determination

- Conducted on all works at time of ingest and when records are modified
 - Public domain worldwide
 - US works published before 1923, US federal government publications, non-US works published prior to 1873
 - Public domain in the United States
 - Non-US works published prior to 1923



Manual Rights Determination

- IMLS-funded CRMS project
 - CRMS-US
 - 2008: US-published works 1923-1963
 - Staff at 4 partner institutions
 - CRMS-World
 - 2011: Expanded to non-US works
 - Staff at 16 partner institutions
 - Double review with additional expert review for conflicts
 - Compliance with copyright formalities
 - As of January 2013 241,541 reviewed, more than 132,644 opened
- Rights Holder Permissions



Rights Database

• System of Precedence

Manual

Bibliographic (automatic)



Lawful uses

- Users who have print disabilities
 - All in-copyright works in HathiTrust currently owned (or owned previously) by the partner institution
 - Must be authenticated
 - Must be on U.S. soil
 - One simultaneous access per copy owned
 - http://www.hathitrust.org/accessibility



Lawful uses (2)

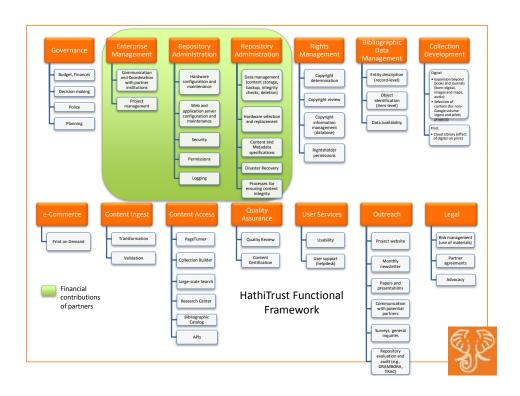
- Out of print and brittle, missing
 - Works must be currently owned (or owned previously) by the partner institution
 - Must be authenticated or accessing work from library premises
 - Must be on U.S. soil
 - One simultaneous access per copy owned
 - http://www.hathitrust.org/out-of-print-brittle
- Access and use statements
 - http://www.hathitrust.org/access use

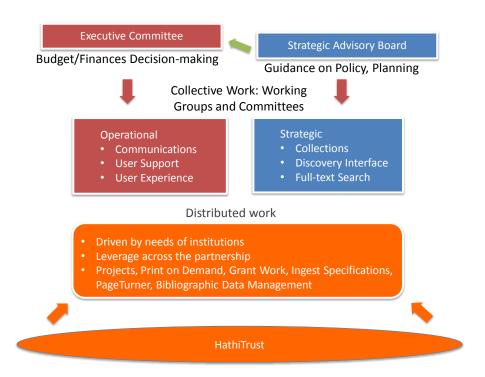


Outline

- The Big Idea 🗸
 - Mission and Goals
- What we're doing to get there
 - Repository and Content
 - Making content available
 - Organizational structure
- How HathiTrust can change the way we work



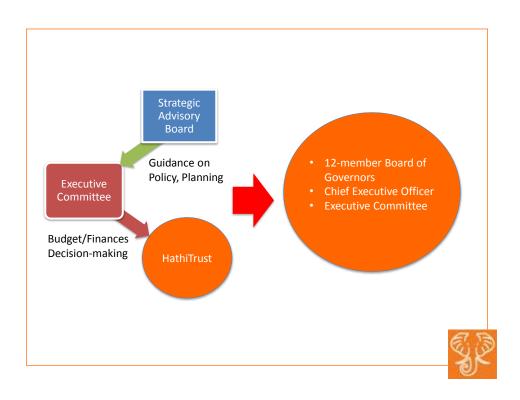




Constitutional Convention

- October 2011
- 52 partners
- 3-year review overseen by SAB
- Ballot Proposals
 - Print monograph storage
 - Approval Process for development initiatives
 - U.S. Government Documents
 - Fee-for-service content deposit
 - Governance





Governance

- Efficient, practical
- Inclusive, collective



Outline

- The Big Idea 🗸
 - Mission and Goals
- What we're doing to get there
 - Repository and Content
 - Making content available
 - Organizational structure
- How HathiTrust can change the way we work

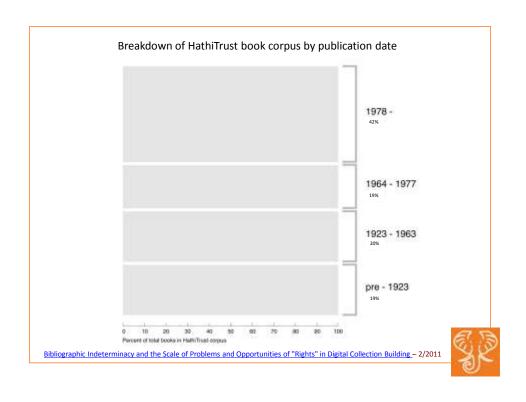


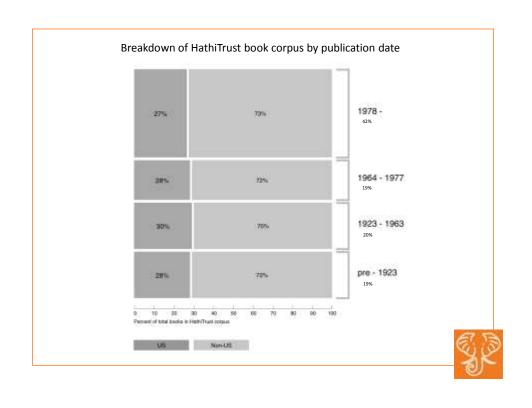
How HathiTrust Can Change the Way We Work

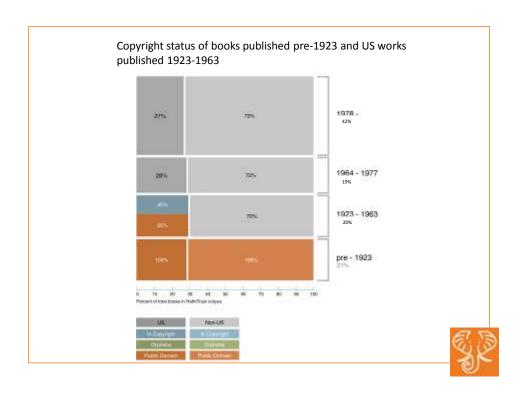


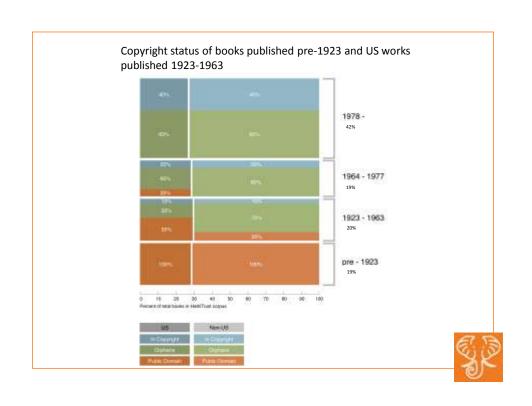
Seeing collective problems as collective

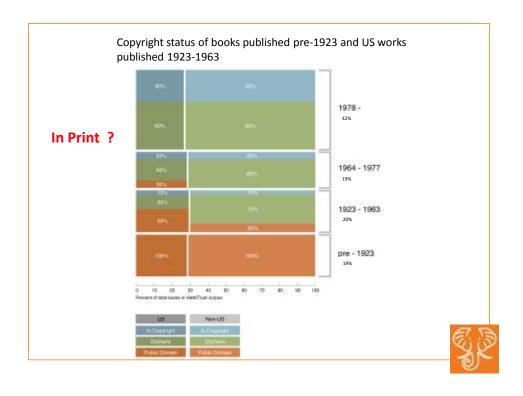












Relationships

- Identification
- Description
- Rights



Relationships

- Identification
- Description
- Rights
- Relationships
 - Bibliographic records



Relationships

- Identification
- Description
- Rights
- Relationships
 - Bibliographic records
 - Bib records and objects



Relationships

- Identification
- Description
- Rights
- Relationships
 - Bibliographic records
 - Bib records and objects
 - Digital objects



Relationships

- Identification
- Description
- Rights
- Relationships
 - Bibliographic records
 - Bib records and objects
 - Digital objects
 - Digital and print

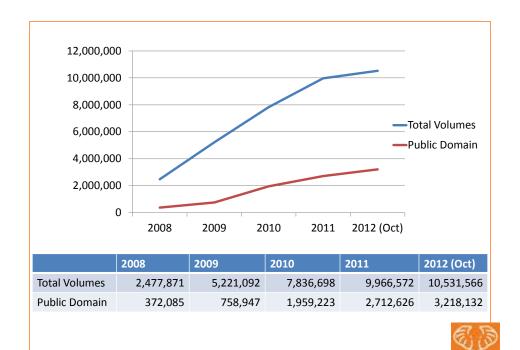


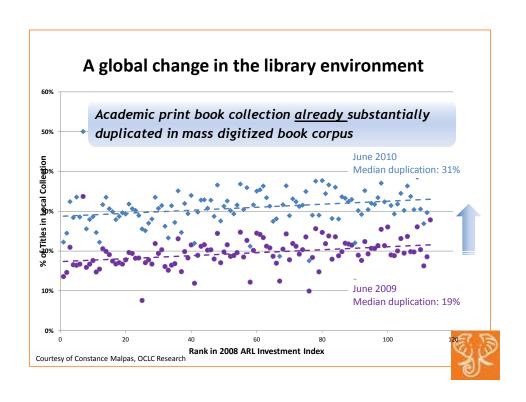
Understanding the relationship between the collective and local

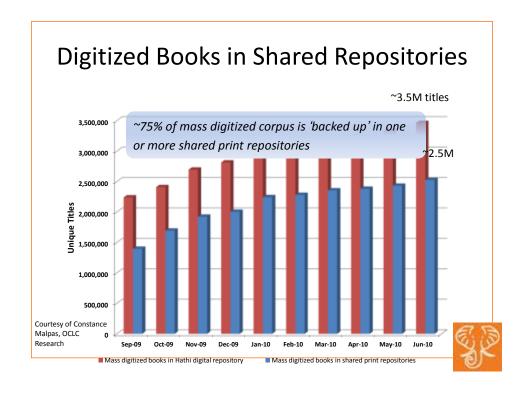


1st model: Price per GB





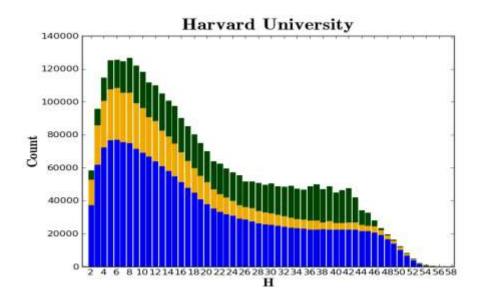


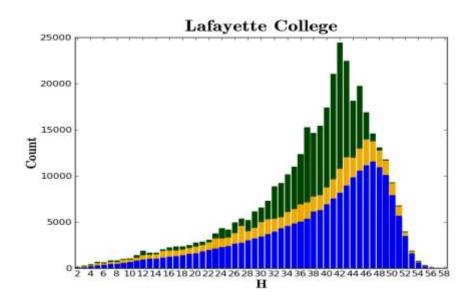


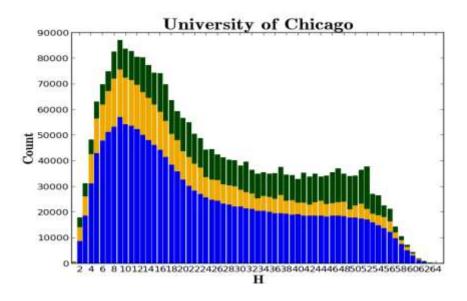
Collection Overlap

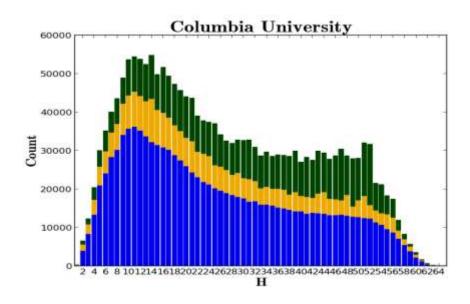
- More than 50% median overlap with ARL institutions; higher for small liberal arts colleges
- New Pricing model based on Print holdings
 - http://www.hathitrust.org/cost
 - Requires print holdings database
 - Also support expansion of legal uses, efforts in deduplication
 - Facilitate individual and collaborative collection development and management operations
- Print monographs archiving

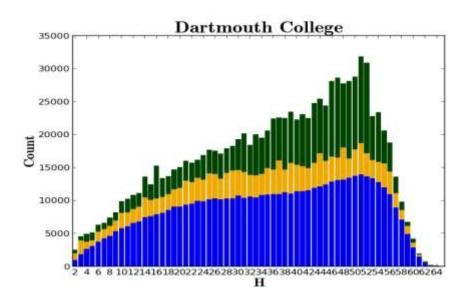


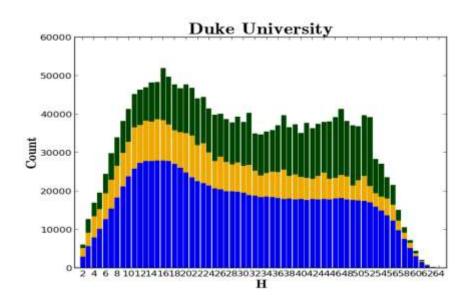


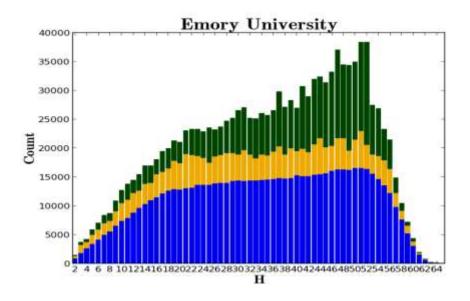


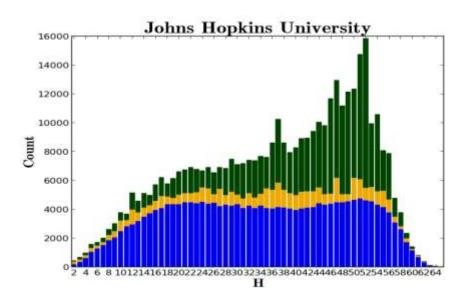












Sourcing and Scaling

http://orweblog.oclc.org/archives/002058.html



- Scale
 - Institution-scale
 - Group-scale
 - Web-scale



- Sourcing
 - Institutional
 - Collaborative
 - Third-party



A new kind of library



Thank you!



How to find out more

- About: http://www.hathitrust.org/about
- Twitter: http://twitter.com/hathitrust
- Facebook: http://www.facebook.com/hathitrust
- Monthly newsletter:
 - <u>http://www.hathitrust.org/updates</u>
 - RSS http://www.hathitrust.org/updates rss
- Contact us: <u>feedback@issues.hathitrust.org</u>
- Blogs: http://www.hathitrust.org/blogs
 - Large-scale Search
 - Perspectives from HathiTrust

