HATHITRUST
A Shared Digital Repository

HathiTrust Data In Detail

September 11, 2012
Jeremy York
Project Librarian, HathiTrust

Information sheet

- http://www.hathitrust.org/documents/hathitrust-data-api-web-client.pdf

Outline

- Content and Metadata

- Repository Organization

- Data availability
  - Rights and agreements

Content and Metadata

Content

- Books and journals
  - Pilots around images, audio, born-digital
- Digitization sources
  - 24 Institutions
  - Google (96.8%, 10,162,104)
  - Internet Archive (2.9%, 301,972)
 - Local (0.3%, 31,840)

Digitization Sources

| Id	Name	Description
1	google	Google
2	lit-dlps-dc	Library IT, DLPS, DC
3	ump	University of Michigan Press
4	ia	Internet Archive
5	yale	Yale University
6	umn	University of Minnesota
7	mhs	Minnesota Historical Society
8	usup	Utah State University Press
9	ucm	Universidad Complutense de Madrid
10	purd	Purdue University
11	getty	Getty Research Institute
12	um-dc-mp	University of Michigan, Duderstadt Center, Millenium Project

Content Sources
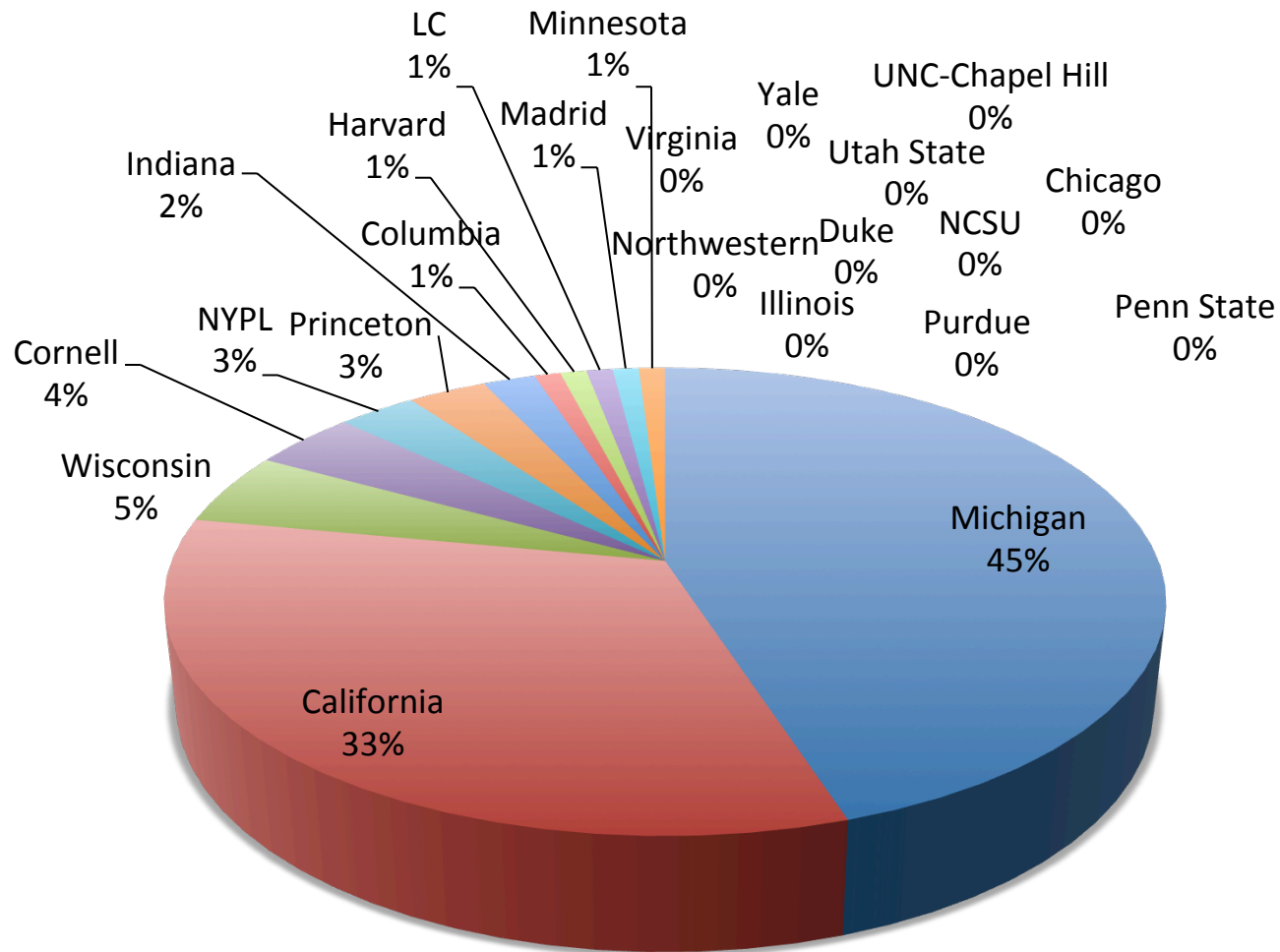
Content

- Largely uniform in technical characteristics
- 3 formats
  - ITU G4 TIFF
  - JP2
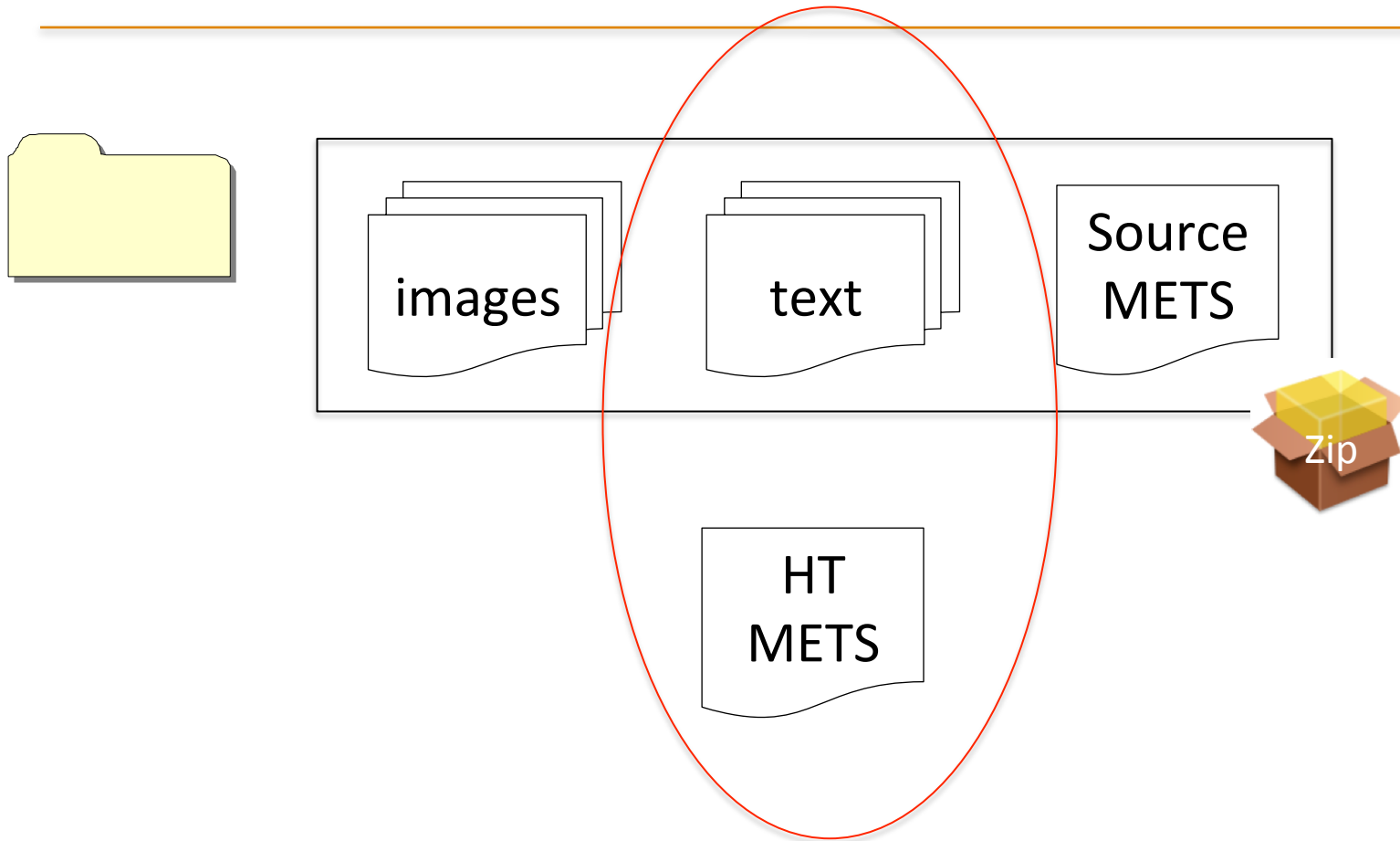 - Unicode (with and without coordinates)

Content Package
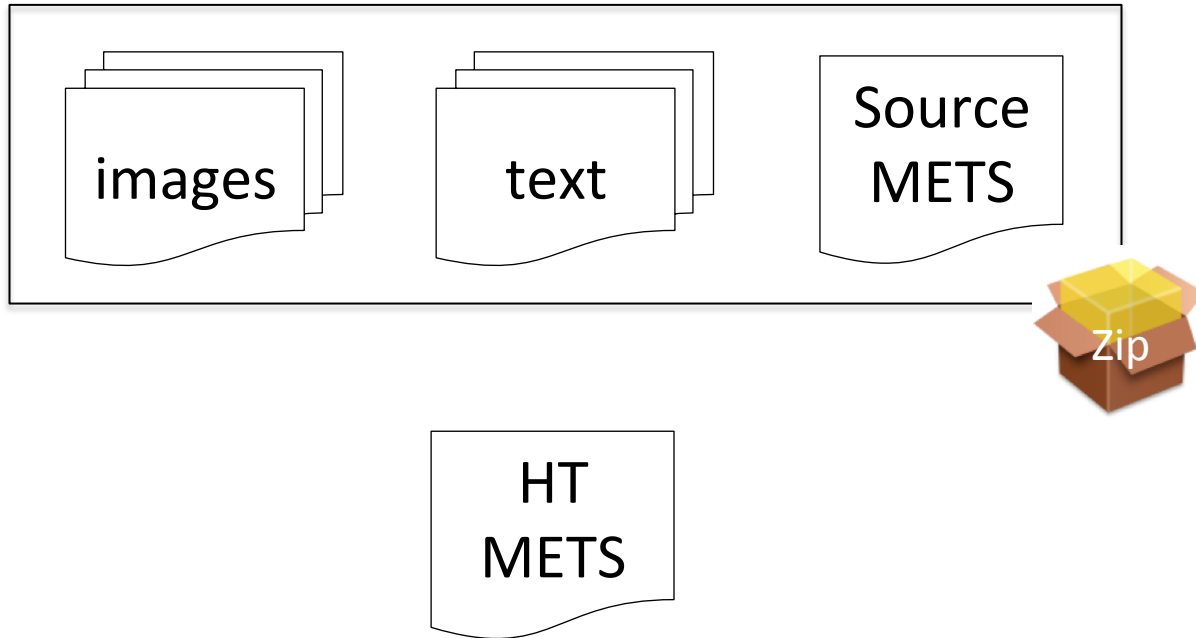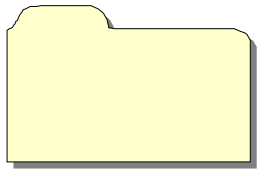
Examples

- Images
- OCR
- Coordinate OCR

Dataset Content Package

images

text

Source METS

HT METS

Zip

What is METS?

- Metadata Encoding and Transmission Standard

- Administrative (including preservation), Technical, and Structural metadata

Why METS

- Can serve as Archival Information Package and a Dissemination Information Package

- Designed to record the relationship between pieces of complex digital objects

- Can be created automatically as texts are loaded or reloaded
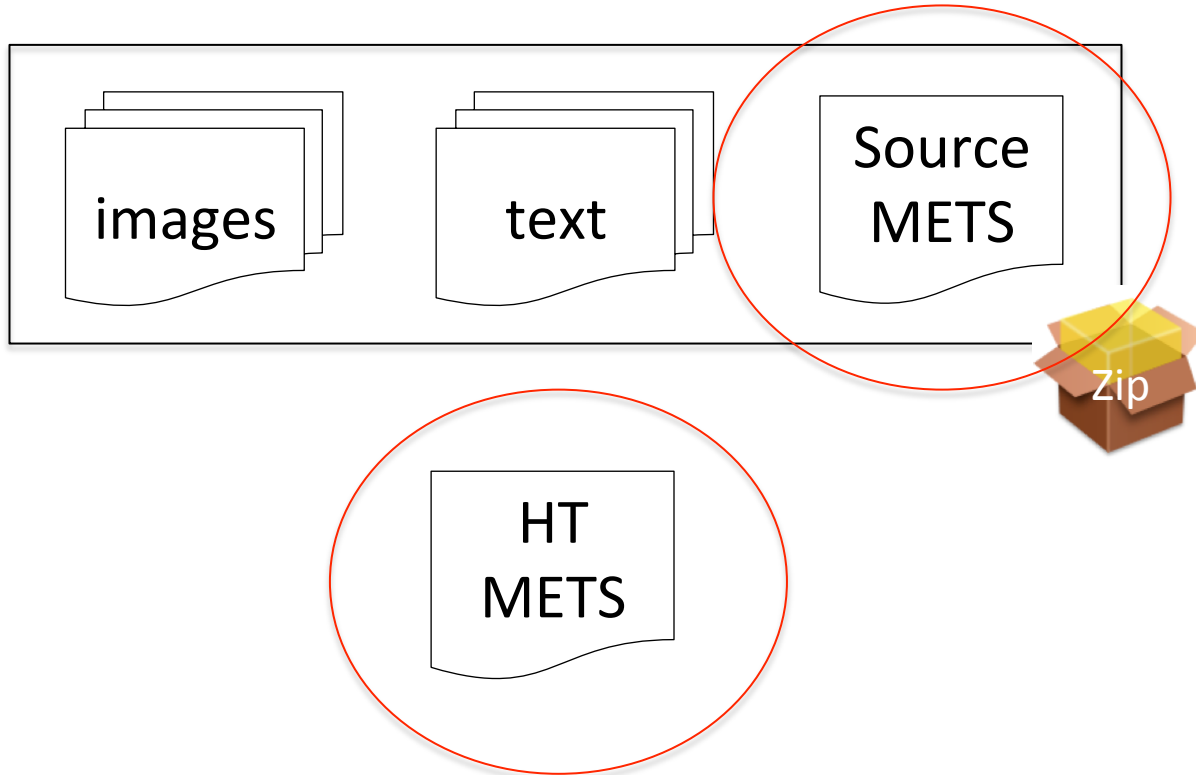
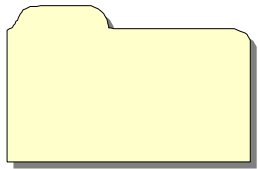- Preservation actions (PREMIS)

Metadata Framework

- Details and specifications at repository level
  - Object specifications / Validation criteria
  - Page-tagging
- Variations at object level
  - Files missing
  - Non-valid files
  - Incorrect file checksums

http://www.hathitrust.org/digital_object_specifications

Content Package

Source METS (1)

- Record of objects prior to ingest into HathiTrust

- Information valuable for preservation or archaeology, but subjective (descriptive, e.g., bibliographic data, page-tags), idiosyncratic, or use not clear.

- "Parking lot" for information we are getting that may be useful in the future.

Source METS (2)

- What's there?
  - dmdSec(s)
  - amdSec
    - Technical and preservation metadata
  - fileSec (images, coordOCR, OCR, …)
    - Mime Type, checksums, file size
  - Physical structMap tying together files with metadata (pg. numbers and features)

HathiTrust METS (1)

- Active record Regularized information generally applicable across the repository
  - Not specific to a particular source
  - Current or near-term use
- Information fundamentally valuable for understanding or using the preserved object in preservation activities after deposit, or in the access and display environments, including the APIs.

HathiTrust METS (2)

- What's there?
  - dmdSec(s)
  - amdSec
  - fileSec with 4 fileGrps (zip, images, OCR, coordOCR)
    - Mime Type, checksums, file size
  - Physical structMap tying together files with metadata (pg. numbers and features)
- HathiTrust METS Profile

Page Feature Mapping (Google)

	A	B	C	D
1	**All Page Tags**	**Source**	**HathiTrust Mapping**	**PageTurner Mapping**
2	FRONT_COVER - Image of the front cover (if the cover of the book was scanned)	Google	FRONT_COVER	Front Cover
3	TITLE_PARTS - Half title page (a sort of preliminary title page before the real one)	Google		
4	TITLE - Title page recto (the front of the real title page)	Google	TITLE	Title Page
5	COPYRIGHT - Title page verso (the back of the real title page)	Google	COPYRIGHT	Copyright
6	PREFACE - First page of each section that appears between the title page verso and the first regularly numbered page. For example, a one-page dedication on page xvi would get this tag, and then the first page of a three-page preface starting on page xviii would also get this. However, if the introduction of the text starts on page 1 (or on an unnumbered page followed by page 2), do not use this tag. Use for components occurring before and after the table of contents.	Google		
7	TABLE_OF_CONTENTS - First page of the table of contents	Google	TABLE_OF_CONTENTS	Table of Contents
8	FIRST_CONTENT_CHAPTER_START - First page of the first chapter with regular page numbering. If the first chapter with regular numbering is called the introduction, that's okay.	Google	FIRST_CONTENT_CHAPT	Section
9	CHAPTER_START - Subsequent chapters with regular page numbering after the first. Also use this for the beginning of each appendix.	Google	CHAPTER_START	Section
10	BACK_COVER - Image of the back cover (if the cover of the book was scanned)	Google		
11	REFERENCES - The first page in a sequence containing endnotes or a bibliography	Google	REFERENCES	Bibliography
12	INDEX - The first page in a sequence containing an index	Google	INDEX	Index
13	BLANK - An intentionally blank page. Don't go looking for these, but if you see one.	Google		

Pagetag Mapping (IA)

	A	B	C	D
1	**All Page Tags**	**Source**	**HathiTrust Mapping**	**PageTurner Mapping**
18	Blank Tissue	Internet Archive	BLANK	?
19	Chapter	Internet Archive	CHAPTER_START	Section
20	Color Card	Internet Archive		
21	Contents	Internet Archive	TABLE_OF_CONTENTS	Table of Contents
22	Copyright	Internet Archive	COPYRIGHT	Copyright
23	Cover	Internet Archive	COVER	?
24	Delete	Internet Archive		Not mapped
25	Foldout	Internet Archive	FOLDOUT	?
26	Footer	Internet Archive		
27	Header	Internet Archive		
28	Illustration	Internet Archive	IMAGE_ON_PAGE	?
29	Illustrations	Internet Archive	IMAGE_ON_PAGE	?
30	Index	Internet Archive	INDEX	Index
31	Norl	Internet Archive		
32	Norma	Internet Archive		
33	normal	Internet Archive		
34	Normal	Internet Archive		
35	Publish Note	Internet Archive		Not mapped
36	Tissue	Internet Archive	BLANK	?
37	Title	Internet Archive	TITLE	Title Page
38	Title Page	Internet Archive	TITLE	Title Page
39	White	Internet Archive		

Pagetag Mapping (DLPS)

	A	B	C	D
1	**All Page Tags**	**Source**	**HathiTrust Mapping**	**PageTurner Mapping**
41	1STPG	DLPS		First Page
42	ACK	DLPS		
43	ADV	DLPS		
44	APP	DLPS	APPENDIX	Appendix
45	BIB	DLPS	BIBLIOGRAPHY	Bibliography
46	BLP	DLPS		
47	CTP	DLPS	TITLE	Title Page
48				
49	DIG	DLPS		Digest
50	FNT	DLPS		
51	HIS	DLPS		
52	IND	DLPS	INDEX	Index
53				
54	LOI	DLPS	LIST_OF_ILLUSTRATIONS	List of Illustrations
55	LOT	DLPS	LIST_OF_MAPS	List of Tables
56	MAP	DLPS		
57	MIS	DLPS		
58	MSS	DLPS		
59	NOT	DLPS	NOTES	Notes

Repository Organization

Source

Ingest

- Bibliographic Data
- Content Package

Data Management

- Bib Data
- Rights Data
- Holdings Data

Storage

- Indiana
- Michigan

Access

- Catalog
- Full-text Search
- PageTurner
- Collections
- APIs
- Datasets
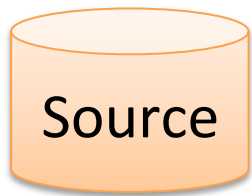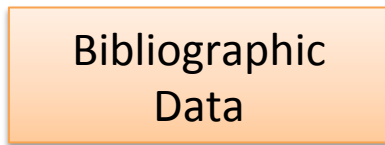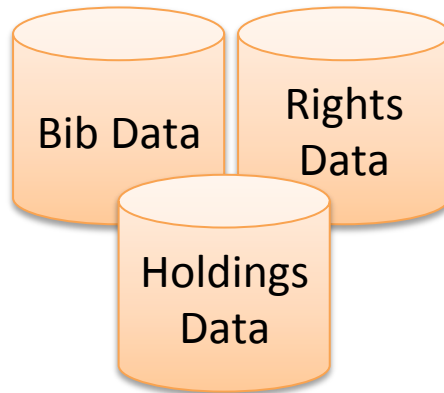
File System
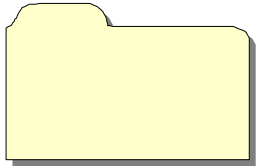
../**uc1**/pairtree_root/b3/54/34/86/b34543486

b34543486.zip

images   text   Source METS

b34543486.mets.xml

HT METS
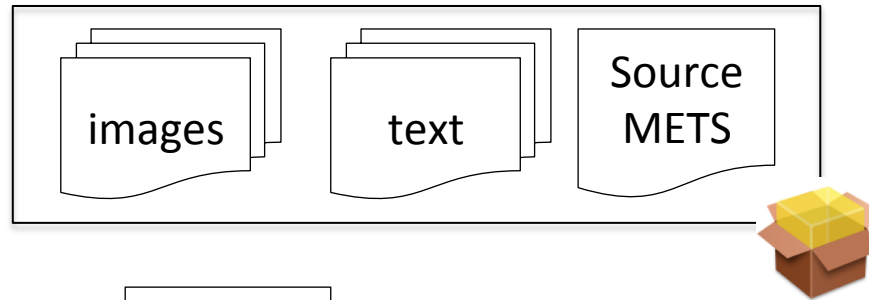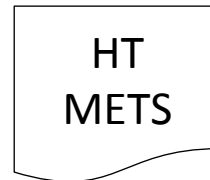
Namespaces

- Namespace
  - 1-4 alphanumeric chars; selected by institution
  - Delineates contributor and unique identifier scheme
  - Example IDs:

  mdp.39015037375253   uc1.b34543486
  miua.aaj0523.1950.001   uc2.ark:/1390/t26973133

Institution	Namespace	Institution	Namespace
Boston College	bc	Universidad Complutense de Madrid	ucm
Columbia University	nnc1, nnc2	University of California	uc1, uc2
Cornell University	coo	University of Chicago	chi
Duke University	dul1	University of Illinois	uiuo, uiug
Harvard University	hvd	University of Michigan	mdp, miua, miun
Indiana University	inu		
Library of Congress	loc	University of Minnesota	umn
New York Public Library	nyp	Minnesota Digital Library	mdl
North Carolina State University	ncs1		
Northwestern University	ien	UNC, Chapel Hill	nc01
Pennsylvania State University	pst	University of Pittsburgh	pitt
	psia	University of Virginia	uva
Princeton University	njp	University of Wisconsin	wu
Purdue University	pur1	Utah State University	usu
	pur2	Yale University	yale

Identifiers

- Prefer identifier used for original object
  - Often barcode
  - Good identifier properties
    - Guaranteed uniqueness
    - Deterministic process for creating new identifiers
    - Internal check scheme
    - Accurate correlation or no correlation to existing names or characteristics (no implied relationships)
  - Facilitate reference
  - Avoid mapping to HathiTrust-generated IDs

Identifier Examples

- mdp.39015037375253

- miua.aaj0523.1950.001

- uc1.b34543486

- uc2.ark:/1390/t26973133

- ucm.5329487288

Google-digitized

chi - University of Chicago
coo - Cornell
hvd - Harvard
ien - Northwester
inu - Indiana University
mdp - University of Michigan
njp - Princeton
nnc1 - Columbia
nyp - NYPL
pst - Penn State
pur1 - Purdue
uc1 - University of California
ucm - Madrid
uiug - University of Illinois
umn - University of Minnesota
uva - University of Virginia
wu - University of Wisconsin

IA-digitized

bc - Boston College
dul1 - Duke
loc - Library of Congress
nc01 - UNC - Chapel Hill
ncs1 - North Carolina State
nnc2 - Columbia
psia - Penn State
uc2 - University of California
uiuo - University of Illinois

Locally-digitized

miua - Michigan
miun - Michigan
mdp - Michigan
ucm - Madrid
usu - Utah State
yale - Yale

Google-digitized

chi - University of Chicago
coo - Cornell
hvd - Harvard
ien - Northwester
inu - Indiana University
mdp - University of Michigan
njp - Princeton
nnc1 - Columbia
nyp - NYPL
pst - Penn State
pur1 - Purdue
uc1 - University of California
ucm - Madrid
uiug - University of Illinois
umn - University of Minnesota
uva - University of Virginia
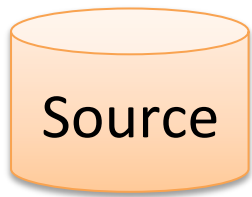wu - University of Wisconsin

IA-digitized

bc - Boston College
dul1 - Duke
loc - Library of Congress
nc01 - UNC - Chapel Hill
ncs1 - North Carolina State
nnc2 - Columbia
psia - Penn State
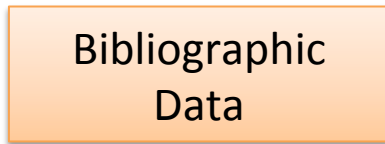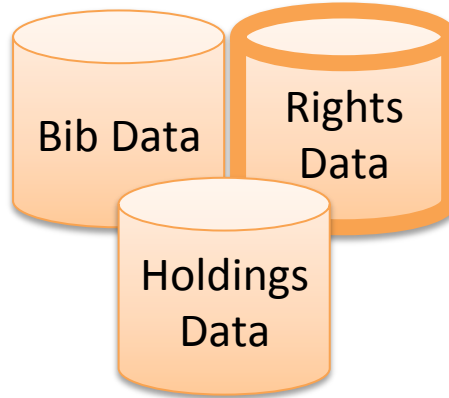uc2 - University of California
uiuo - University of Illinois

Locally-digitized

miua - Michigan
miun - Michigan
mdp - Michigan
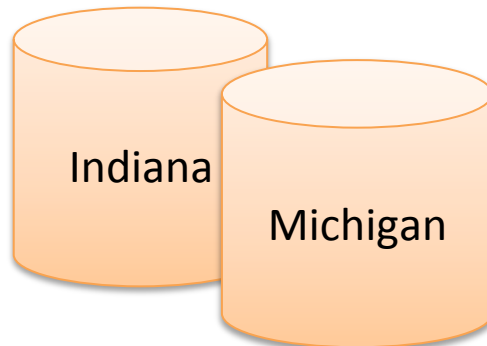**ucm - Madrid**
usu - Utah State
yale - Yale

Copyright

- Bibliographic metadata
- Automatic and manual rights determination

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 1872
  - Public domain in the United States
   - Non-US works published prior to 1923

Manual Rights Determination

- IMLS-funded CRMS project
  - US-published works 1923-1963
  - Conformance with formalities
  - Expanding to non-US works
  - Double-blind review with expert review for conflicts
  - Staff at 4 HathiTrust partner institutions (15 will take part in non-US)
  - As of February 2012 ~190,000 reviewed, more than 100,000 opened

- Rights Holder Permissions

# Rights Database

- System of Precedence

| Manual |
|:------:|

↑

Bibliographic (automatic)
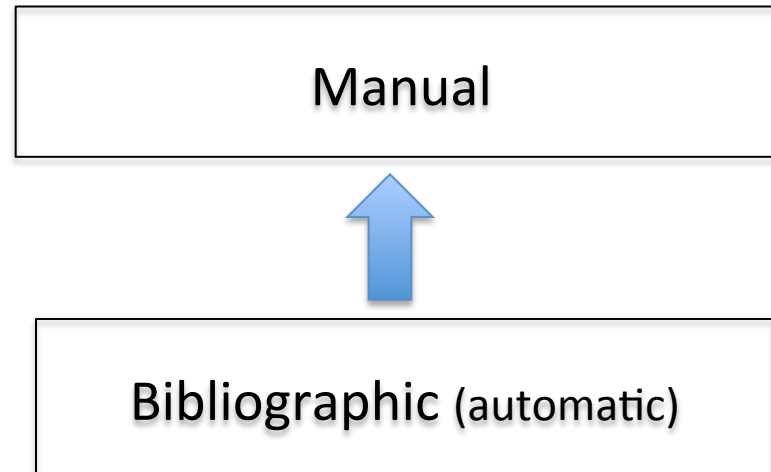
Rights Attributes

| id	name	type	dscr
1	pd	copyright	public domain
2	ic	copyright	in-copyright
3	opb	copyright	out-of-print and brittle (implies in-copyright)
4	orph	copyright	copyright-orphaned (implies in-copyright)
5	und	copyright	undetermined copyright status
6	umall	access	available to UM affiliates and walk-in patrons (all campuses)
7	world	access	available to everyone in the world
8	nobody	access	available to nobody; blocked for all users
9	pdus	copyright	public domain only when viewed in the US
10	cc-by	copyright	Creative Commons Attribution
11	cc-by-nd	copyright	Creative Commons Attribution-NoDerivatives
12	cc-by-nc-nd	copyright	Creative Commons Attribution-NonCommercial-NoDerivatives
13	cc-by-nc	copyright	Creative Commons Attribution-NonCommercial
14	cc-by-nc-sa	copyright	Creative Commons Attribution-NonCommercial-ShareAlike
15	cc-by-sa	copyright	Creative Commons Attribution-ShareAlike
16	orphcand	copyright	orphan candidate - in 90-day holding period (implies in-copyright)
17	cc-zero	copyright	Creative Commons Zero license (implies pd)
18	und-world	copyright	Undetermined copyright status and permitted as world-viewable by the depositor
19	Ic-us	copyright	In copyright in the US

Rights Determination Reason Codes

id	name	dscr
1	bib	bibliographically-derived by automatic processes
2	ncn	no printed copyright notice
3	con	contractual agreement with copyright holder on file
4	ddd	due diligence documentation on file
5	man	manual access control override; see note for details
6	pvt	private personal information visible
7	ren	copyright renewal research was conducted
8	nfi	needs further investigation (copyright research partially complete; an ambiguous, unclear, or other time-consuming situation was encountered)
9	cdpp	title page or verso contain copyright date and/or place of publication information not in bib record
10	cip	condition review and in-print status research was conducted
11	unp	unpublished work
12	gfv	Google viewability set at VIEW_FULL
13	crms	derived from multiple reviews in the Copyright Review Management System (CRMS) via an internal resolution policy; consult CRMS records for details
14	add	author death date research was conducted or notification was received from authoritative source
15	exp	expiration of copyright term for non-US work with corporate author
16	Del	Deleted from repository; see note for details
17	Gatt	Non-US public domain work restored to in-copyright in the US by GATT
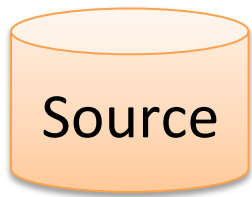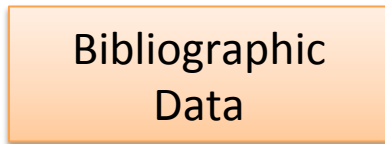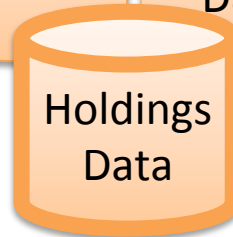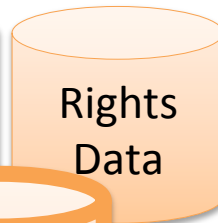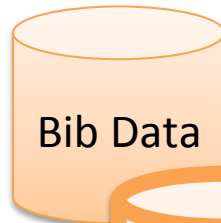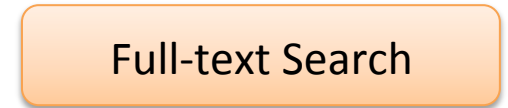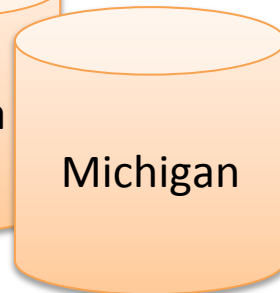
Holdings Database

- Single-part monographs
  - OCLC #; Local system ID; Timestamp; Holding Status; Condition

- Multi-part monographs
  - Include enumeration and chronology

- Serials
  - OCLC #; Local system ID; Timestamp; ISSN

Harvard University
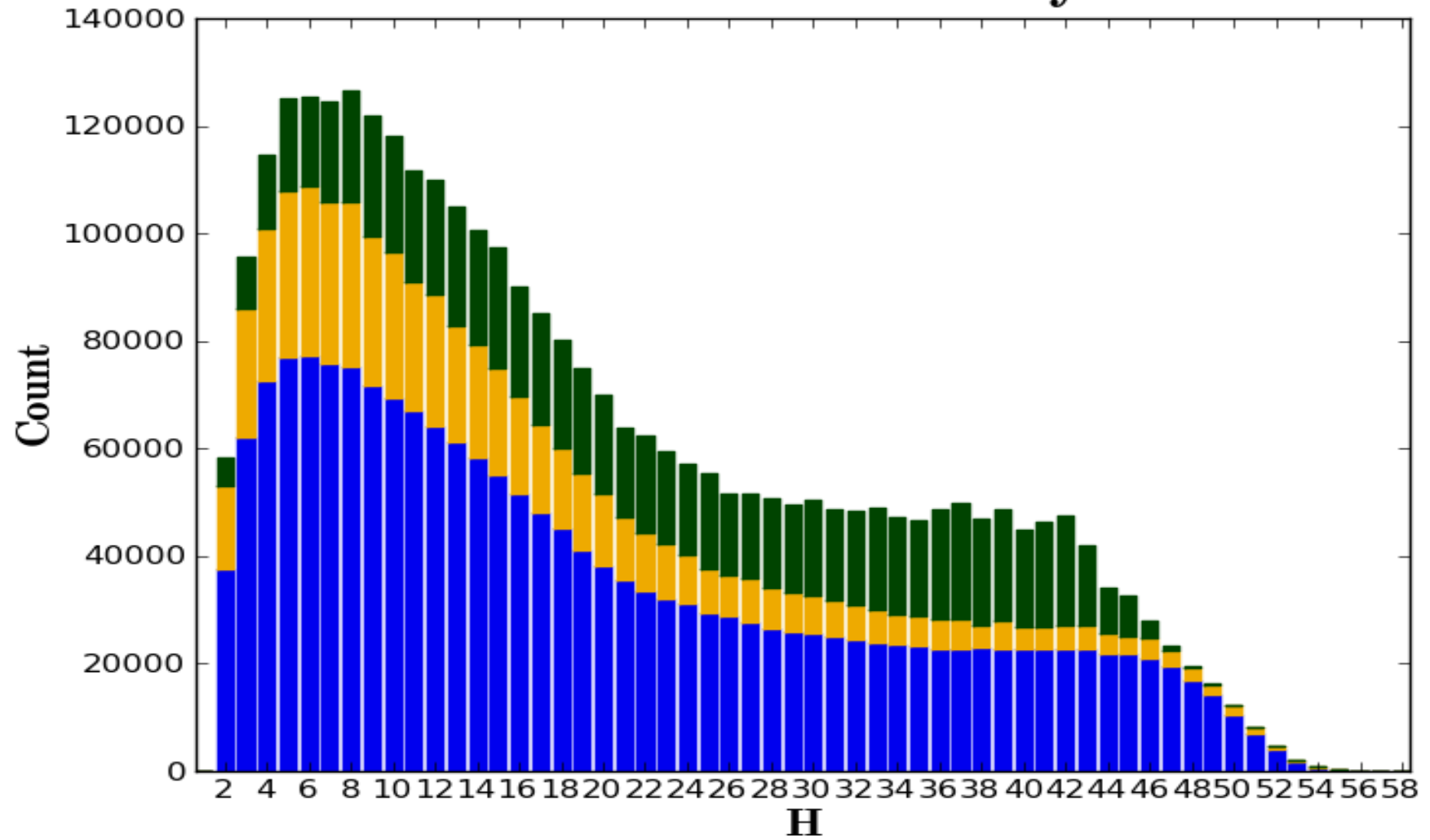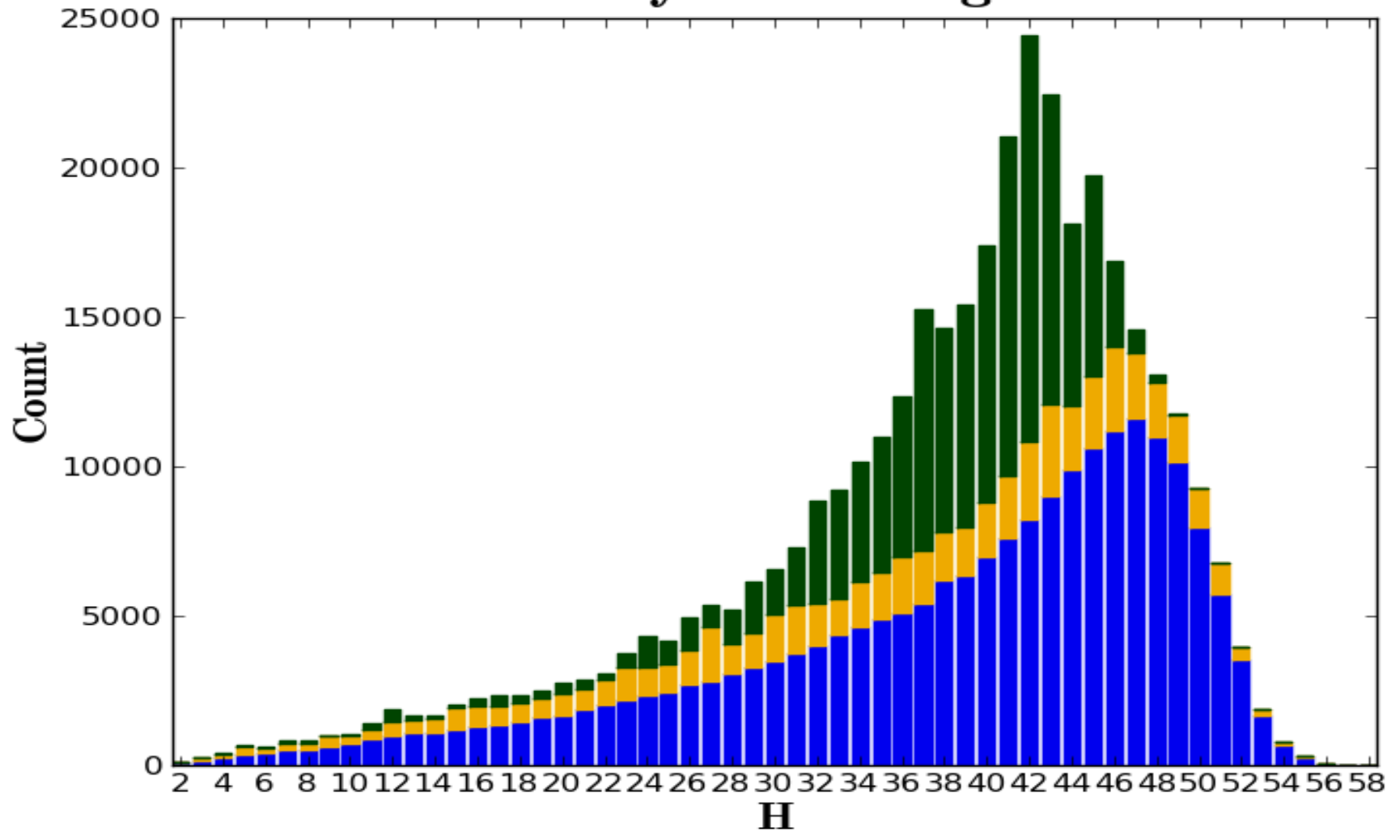
Data Availability

What is available?

images   text   Source METS

Zip

HT METS

- Bibliographic metadata
- Rights metadata

How is it available?

- Web interfaces
- APIs
  - Data API
  - Bib API
- Data feeds and distribution
  - Hathifiles
  - OAI
  - Datasets

ⓘ Catalog Search

Search information **about** the items.

[   ] [Find]

[All Fields   ▾] ☐ Full view only

Advanced Catalog Search | Search Tips

📖 Full-text Search

Search words that occur **within** the items.

[   ] [Find]

☐ Full view only

Advanced Full-text Search | Search Tips

📑 Collections

Browse, search, or make HathiTrust collections.

[View Public Collections]

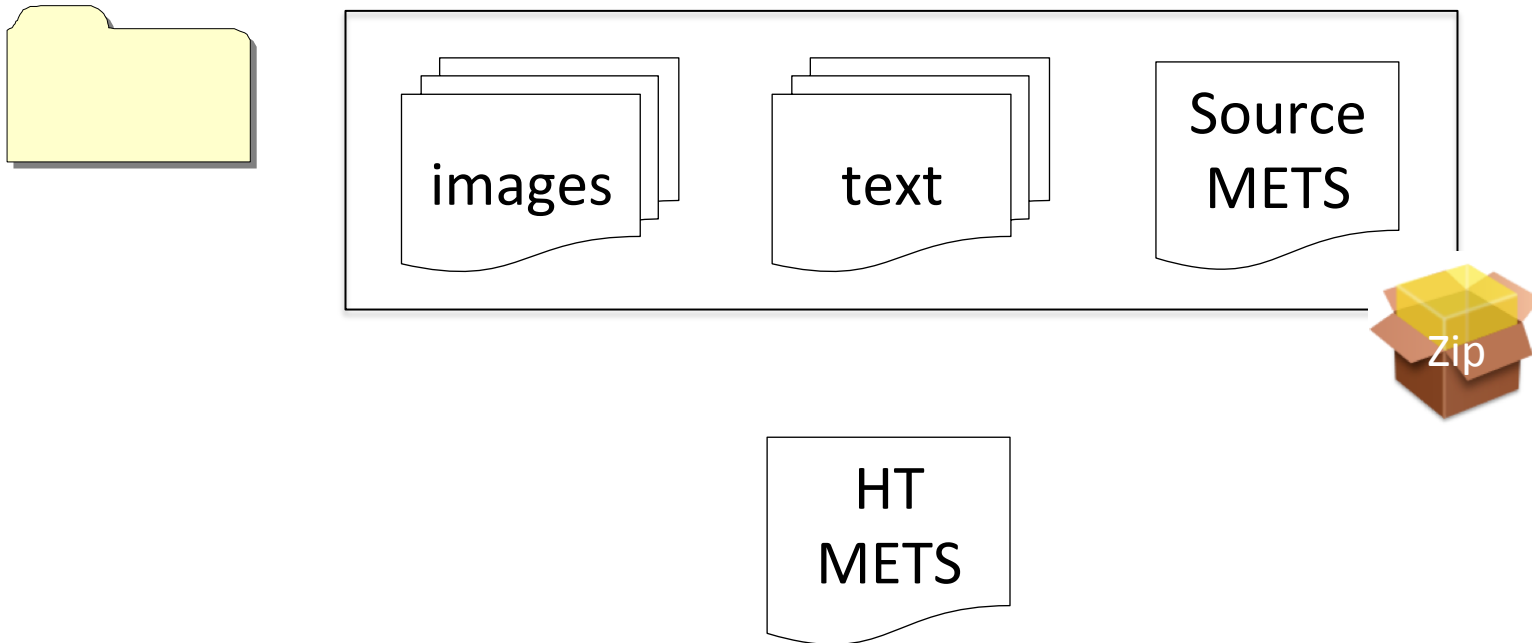Collections are a way to group items for private or public use. The full-text of items within a collection can be searched independently of the full digital library. Read more about creating collections.

Recent News and Publications

Board of Governors Election Results

Try our new mobile website!

Information about the Authors Guild Lawsuit

- Update on July 2012 Activities
- Getting everything you want from HathiTrust - The Stone and the Shell
- HathiTrust And TRAC
- HathiTrust: On TRAC
- Update on June 2012 Activities

view more news ››

Currently Digitized

10,497,399 total volumes
5,533,530 book titles
273,684 serial titles
3,674,089,650 pages
471 terabytes
124 miles
8,529 tons
3,189,660 volumes (~30% of total) in the public domain
View visualizations of HathiTrust
call numbers, languages, and dates

statistics information ››

Featured Collection

Utah State University Press

The Utah State University Press titles available for full view in HathiTrust.

HATHI TRUST Digital Library

About   Help   Feedback

Search   Public Collections   My Collections

Catalog Search

ⓘ **Catalog Search**   library of universal knowledge   All Fields ▾   ☐ Full view only   Find   **Advanced Catalog Search**   *Search Tips*

Narrow Search

Viewability
Limited (search-only) (21355)
Full view (3012)

Subject
United States (483)
Great Britain (212)
Knowledge, Theory of (202)
Knowledge, Theory of. (181)
Bible. (161)
more...

Author
Brüll, Ignaz, 1846-1907. (53)
University of Michigan. Library. (50)
Aristotle. (44)
Harvard University. Library. (37)
Cicero, Marcus Tullius. (31)
more...

Language
English (20052)
German (902)
French (755)
Italian (311)

✉ **Email this Search**

Showing **1 - 20** of 24000 Results for **library of universal knowledge**

Sort   Relevance ▾

Relevance
Date (newest first)
Date (oldest first)
Author

1   2   3   4   5   6   7   8   9   10   11   Next »   [1200]

Library of universal knowledge. A reprint of the last (1880) Edinburgh and London edition of Chambers' encyclopaedia, with copious additions by American editors.
Published 1880
Multiple volumes

CUU ... conference proceedings.
by Conference on Universal Usability.
Published 2000
Multiple volumes

Library bulletin of abstracts.
by Universal Oil Products Company. Research Laboratories.
Published 1926
Multiple volumes

Proceedings of the ... International Conference on Knowledge Capture.

HATHI TRUST Digital Library

Search   Public Collections   My Collections

Catalog Search

Catalog Search   [   ] [All Fields ▼] ☐ Full view only (Find)   Advanced Catalog Search   Search Tips

Similar Items

Chamber's encyclopaedia : a dictionary of universal knowledge for the people.
Published: (1875)

Chambers' encyclopaedia. A dictionary of universal knowledge.
Published: (1888)

Chambers's encyclopaedia; a dictionary of universal knowledge.
Published: (1923)

Chambers's encyclopaedia; a dictionary of universal knowledge for the people,
Published: (1860)

Check list for American local history, Reprinted with additions from the Bulletins of the Boston public library,
By: Perkins, Frederic B. 1828-1899.
Published: (1876)

Westminster drolleries, both parts, of 1671, 1672; being a choice collection of songs and poems, sung at court & theatres: with additions made by 'A person of quality.' Now first reprinted from the original editions.
Published: (1875)

Roads & highways: communicated to the editor of The Edinburgh encyclopaedia,
By: Stevenson, Robert, 1772-1850.
Published: (1824)

The library of choice literature and encyclopædia of universal authorship ...
By: Spofford, Ainsworth Rand, 1825-1908.
Published: (1893)

Boswell's correspondence with the Honourable Andrew Erskine, and his Journal of a tour to Corsica (reprinted from the original editions).
By: Boswell, James, 1740-1795.

Showing record 1 of 24000 Next record ➡

Email this   Cite this   Export to Endnote

Library of universal knowledge. A reprint of the last (1880) Edinburgh and London edition of Chambers' encyclopaedia, with copious additions by American editors.

Language(s):	English
Published:	New York : American Book Exchange, 1880-81.
Subjects:	Encyclopedias and dictionaries.
Physical Description:	15v. illus.
Original Format:	Book Encyclopedias
Locate a Print Version:	Find in a library

Viewability:

☐ Full view v.04 *(original from University of California)*
☐ Full view v.08 *(original from University of California)*

HATHI TRUST Digital Library

Home   About   Collections   My Collections

Full-text Search

Full-text Search

"library of universal knowledge"   ☐ Full view only   [Find]   Advanced Full-text Search

Search Results: 1,176 items found for **"library of universal knowledge"** in *Everything*

[25 per page ▾]   Prev **1** 2 3 4 5 6 7 8 9 10 ... 48   >> Next

Refine Results

Viewability
Full View (832)
Limited (search only) (344)

Subject
United States (132)
American literature (101)
Encyclopedias and dictionaries. (87)
American literature Bibliography Periodicals. (63)
Publishers and publishing (59)
more...

Author
Library of Congress. Copyright Office. (15)
Griffis, William Elliot, 1843-1928. (14)
Dunning, William Archibald, 1857-1922. (10)
Library of Congress. Copyright Office (10)
Allibone, S. Austin 1816-1889. (9)
more...

Language
English (1,120)
German (18)
French (10)
Undetermined (10)
Japanese (6)

Select all on page ☐   [Select Collection ▾]   [Add Selected]

☐ **General encyclopedias in print. 1963**
Published 1963

ⓘ Catalog Record   Limited (search-only)

☐ **History by the topical method.**
by Woody, John W. [from old catalog]
Published 1889

ⓘ Catalog Record   Full view

☐ **Erasmus & Henry VIII. By J.H. Merle d'Aubigné...**
by Merle d'Aubigné, J. H. 1794-1872.
Published 1884

ⓘ Catalog Record   Full view

☐ **General encyclopedias in print. 1964**
Published 1963

ⓘ Catalog Record   Limited (search-only)

☐ **Anglo-American general encyclopedias : a historical bibliography, 1703-1967 / by S. Padraig Walsh.**
by Walsh, S. Padraig, 1922-
Published 1968

ⓘ Catalog Record   Limited (search-only)

☐ Anglo-American general encyclopedias : a historical bibliography, 1703-1967 / by S. Padraig Walsh.

« Back to "Full text search" results

Search in this text [ ] Find

Classic View   New   Try our new views!   Scroll   Flip   Thumbnails   Plain Text

Jump to Section ▾ Go   Jump to [ ] Go

About this Book

History by the topical method. . Woody, John W. [from old catalog]

View full catalog record

Copyright: Public Domain.

Get this Book

Download this page (PDF)

Download whole book (PDF)

Add to Collection

Login to make your personal collections permanent

Select Collection ▾

Add

Share

Permanent link to this book

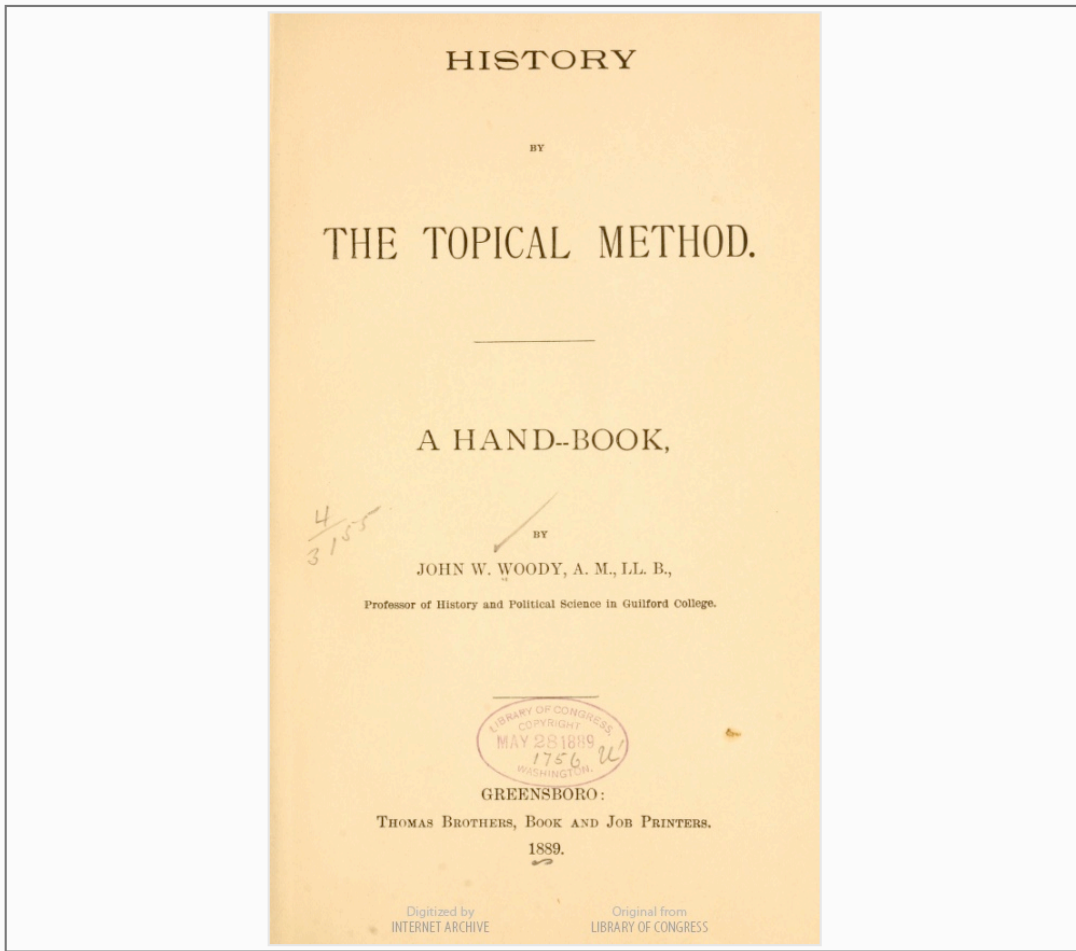http://hdl.handle.net/2027/loc.ark:/13960/t0000h93g

Link to this page

http://hdl.handle.net/2027/loc.ark:/13960/t0000h93g'

Version: 2011-03-12 03:20 UTC ?

HISTORY

BY

THE TOPICAL METHOD.

A HAND--BOOK,

BY

JOHN W. WOODY, A. M., LL. B.,

Professor of History and Political Science in Guilford College.

GREENSBORO:
THOMAS BROTHERS, BOOK AND JOB PRINTERS.
1889.

Jump to Section ▾ Go   Jump to [ ] Go

« Back to page

Search in this text "library of universal knowledge" Find

◀ Go to the beginning of the book

"library of universal knowledge" matched **10** pages in this item.

Viewing results for: 1 to 10 of 10 pages

p.61 - 2 matching terms
…ers's **Library of Universal Knowledge**. I Machune. Mosheim's Ecclesiastical History, J _ Vol. T, 373, 403. I Lippincott's Encyclopsedia of Religious Knowl- I edge. I American Sunday School Union. The Inquisi- tion in Spain and other countries. I Backhouse. Witnesses for Chris…

…ers's **Library of Universal Knowledge**. Motley. The Rise of the Dutch Republic, Vol. II, 84— …

p.36 - 1 matching term
…ers's **Library of Universal Knowledge**. Vol. vn. Lippincott's Pronouncing Biographical Dic- tionary. Smith. Classical Dictionary. _ Barnes. A Brief History of Rome, 148-152. LESSON XXV. Position. Size and shape. Boundaries. 4. Physical features. - 1. Surftice. …

p.43 - 1 matching term
…bers. **Library of Universal Knowledge**. …

p.45 - 1 matching term
…rrs's **Library of Universal Knowledge**. Lippincott's BiogrBphical Dictionary. 7. References. …

p.49 - 1 matching term
…ers's **Library of Universal Knowledge**. Sivinion. Outlines of the World's History. LESSON XLVIL 1. Time 1095 to 1270. 2. Number. 3. Condition of the I Church at the^ beginning of L 4. The social con- fl- dition of Eu- j ^ rope at the j ^' beginning of The firs…

p.50 - 1 matching term
…ers's **Library of Universal Knowledge**. Guizot. The History of France, Vol. I, 205-210. Abbott. Markham's France, 53 — [a. H. Johnson. The Normans in Europe, 35-37, ri. 2. 3. 4. 5. .6 References. ■ …

p.51 - 1 matching term
…ers's **Library of Universal Knowledge**. Ahhott. History of Alfred the Great. …

p.54 - 1 matching term
…ers's **Library of Universal Knowledge**. American Sunday School Union. Caxton and The Art of Printing. Montgomerij. The Leading Facts of English History, 167. Green. History of the English People, Vol. II. 53. 4. Results of. …

p.80 - 1 matching term
…ers's **Library of Universal Knowledge**.) American Cyclopaedia. Ihalheimer. Eclectic History of the United States, 91-92. Quackenbos. History of U. S., 127 — Labberton. Historical Atlas. 10. References. …

p.93 - 1 matching term
…ers's **Library of Universal Knowledge**. Lippincott's Pronouncing Biographical Dic- tionary. In the rule of Cromwell the current of history was effected by the well directed efforts of a strong character. In that of Charles II. by the wreck- less acts of an aimless profligate. Both advanc…

Viewing results for: 1 to 10 of 10 pages

« Back to "Full text search" results

Search in this text [   ] Find

ⓘ About this Book

History by the topical method. . Woody, John W. [from old catalog]

View full catalog record

Copyright: Public Domain.

⬇ **Get this Book**

Download this page (PDF)

Download whole book (PDF)

🗒 **Add to Collection**

Login to make your personal collections permanent

Select Collection   ⇕

Add

⚙ **Share**

Permanent link to this book

http://hdl.handle.net/2027/loc.ark:/13960/t0000h93g

Link to this page

http://hdl.handle.net/2027/loc.ark:/13960/t0000h93g

Version: 2011-03-12 03:20 UTC ❓

Classic View   New   Try our new views!   ▤ Scroll   📖 Flip   ▦ Thumbnails   Plain Text

Jump to Section ⇕ Go   🔍 🔍 ↺ ↻   Jump to [   ] Go ⏮ ◀ ▶ ⏭

HISTORY

BY

THE TOPICAL METHOD.

———

A HAND--BOOK,

HISO^ORY

BY

THE TOPICAL METHOD.

A HAND-BOOK,

BY

JOHN W. WOODY, A. M., LL. B.,

Professor of History and Political Science in Guilford College.

MAY2818B9

GREENSBORO :
Thomas Brothers, Book and Job Printers.

1889.

How is it available?

- Web interfaces ✔
- APIs
  - Data API
  - Bib API
- Data feeds and distribution
  - Hathifiles
  - OAI
  - Datasets

Data API Demonstration

- http://babel.hathitrust.org/cgi/kgs/portal

- Examples

  - mdp.39015071393550 (seq 7)

  - loc.ark:/13960/t0000h93g (seq 7)

- Page Image
- Page OCR
- Page Coordinate OCR
- METS
- Object Metadata
  - Rights, page numbers and features
- Page Metadata
  - Rights, page sequence and number, format

Bib API

- Gives bibliographic, volume, rights information
- When supplied with
  - OCLC, LCCN, LSSN, ISBM, HTID, Record ID
- Returns "brief" and "full" results
  - Full includes MARCXML in JSON wrapper

http://catalog.hathitrust.org/api/volumes/**brief**/<id type>/<id value>.json
http://catalog.hathitrust.org/api/volumes/**full**/<id type>/<id value>.json

Examples: mdp.39015071393550; loc.ark:/13960/t0000h93g

How is it available?

- Web interfaces ✔

- APIs ✔
  - Data API
  - Bib API

- Data feeds and distribution
  - Hathifiles
  - OAI
  - Datasets

OAI

- OAI sets (MARC21 or Dublic Core)
  - Public domain and open access (set=hathitrust:pd)
  - Public domain in the United States (set=hathitrust:pdus)
  - All (PD, OA, PDUS) (set=hathitrust)

http://quod.lib.umich.edu/cgi/o/oai/oai?verb=ListRecords&metadataPrefix=marc21&set=hathitrust

Hathifiles

- Tab-delimited inventory files

- Aggregated monthly

- Daily incremental files

- Contain
  - Identifiers
  - Limited bibliographic information
  - Rights, language, gov docs status information

HATHI TRUST Digital Library

Search About HathiTrust [Go]

About › sites/www.hathitrust.org/files/hathifiles

Hathifiles

Displaying *sites/www.hathitrust.org/files/hathifiles*.

Contains 43 files totaling 1212.95 MB in size.

Name ▲	Size
hathi_full_20120801.txt.gz	582.09 MB
hathi_full_20120901.txt.gz	584.68 MB
hathi_upd_20120801.txt.gz	2.49 MB
hathi_upd_20120802.txt.gz	1.97 MB
hathi_upd_20120803.txt.gz	894.17 KB
hathi_upd_20120804.txt.gz	760.59 KB
hathi_upd_20120805.txt.gz	662.04 KB
hathi_upd_20120806.txt.gz	664.25 KB
hathi_upd_20120807.txt.gz	619.04 KB
hathi_upd_20120808.txt.gz	1.14 MB
hathi_upd_20120809.txt.gz	1.29 MB
hathi_upd_20120810.txt.gz	792.76 KB
hathi_upd_20120811.txt.gz	634.93 KB
hathi_upd_20120812.txt.gz	601.06 KB
hathi_upd_20120813.txt.gz	627.96 KB
hathi_upd_20120814.txt.gz	908.87 KB
hathi_upd_20120815.txt.gz	1.75 MB
hathi_upd_20120816.txt.gz	812.94 KB
hathi_upd_20120817.txt.gz	1.27 MB
hathi_upd_20120818.txt.gz	3.05 MB
hathi_upd_20120819.txt.gz	1.09 MB
hathi_upd_20120820.txt.gz	1.09 MB
hathi_upd_20120821.txt.gz	1.25 MB
hathi_upd_20120822.txt.gz	918.16 KB
hathi_upd_20120823.txt.gz	1.51 MB
hathi_upd_20120824.txt.gz	944.77 KB
hathi_upd_20120825.txt.gz	629.03 KB
hathi_upd_20120826.txt.gz	524.2 KB
hathi_upd_20120827.txt.gz	688.2 KB
hathi_upd_20120828.txt.gz	692.41 KB
hathi_upd_20120829.txt.gz	1.37 MB
hathi_upd_20120830.txt.gz	1.04 MB
hathi_upd_20120831.txt.gz	867.8 KB
hathi_upd_20120901.txt.gz	734.1 KB
hathi_upd_20120902.txt.gz	711.16 KB
hathi_upd_20120903.txt.gz	643.44 KB
hathi_upd_20120904.txt.gz	762.64 KB
hathi_upd_20120905.txt.gz	1.21 MB
hathi_upd_20120906.txt.gz	2.46 MB
hathi_upd_20120907.txt.gz	3.31 MB
hathi_upd_20120908.txt.gz	2.52 MB
hathi_upd_20120909.txt.gz	633.44 KB
hathi_upd_20120910.txt.gz	25.43 KB

Printer-friendly version

- Searching, Reading, and Building Collections
- Technology, Standards, and Specifications
- Getting Content Into HathiTrust
- Data Availability and APIs
  - HathiTrust Bibliographic API
  - HathiTrust Data API
  - Hathifiles
  - Hathifiles Description
  - Datasets
- Policies
- Statistics and Visualizations
- HathiTrust Personas

Data Element	Example
Volume identifier	coo.31924003924275
Access	deny
Rights	ic
University of Michigan Record #	002052896
Enumeration/Chronology	Band I
Source	COO
Source Institution Record #	17132
OCLC numbers	62370740
ISBNs	
ISSNs	gs 12000204
LCCNs	

Data Element	Example
Title	Anleitung zur bestimmung der karbonpflanzen…
Imprint	Kommissionsverlag von Craz & Gerlach (J. Stettner) 1911-
Rights determination reason code	bib
Date of last update	2011-04-11 20:32:41
Government document	0
Publication date	1911
Publication place	gw
Language	ger
Bibliographic format	BK

Datasets

- Non-Google-digitized Dataset (300,000+)
  - PD, PDUS, Open Access
  - Signed researcher statement
- Google-digitized (2.2 million+)
  - PD, PDUS, Open Access
  - Agreement between institution and Google
  - Brief proposal
    - Characterize texts
    - Provide ids (custom sets possible)
    - Research, results, use of results
 - Signed researcher statement

Digitization Sources

| Id	Name	Description
1	google	Google
2	lit-dlps-dc	Library IT, DLPS, DC
3	ump	University of Michigan Press
4	ia	Internet Archive
5	yale	Yale University
6	umn	University of Minnesota
7	mhs	Minnesota Historical Society
8	usup	Utah State University Press
9	ucm	Universidad Complutense de Madrid
10	purd	Purdue University
11	getty	Getty Research Institute
12	um-dc-mp	University of Michigan, Duderstadt Center, Millenium Project

Dataset structure
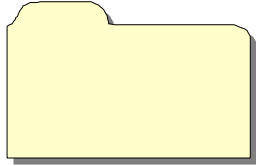
id (list of ids in dataset)
meta.tar.gz (bibliographic data)
loc
mdp
uc1

b34543486.zip
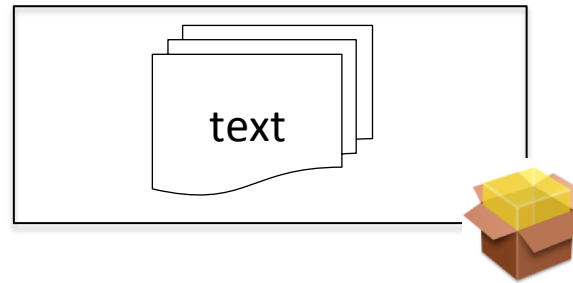
text

b34543486.mets.xml

HT
METS

How is it available?

- Web interfaces ✔
- APIs ✔
  - Data API
  - Bib API
- Data feeds and distribution ✔
  - Hathifiles
  - OAI
  - Datasets

Which Bibliographic Data?

- Bibliographic data from Dataset
  - One record per item; enum/chron as appears in record; dates not normalized
- Bib API
  - Dates at bib level; if no date in Date1 of 008, returns 260|c; can query to determine if multiple copies or items
- Hathifiles
  - Dates extracted per-item; no date information if bib for item has no Date1 in 008 or enum/chron
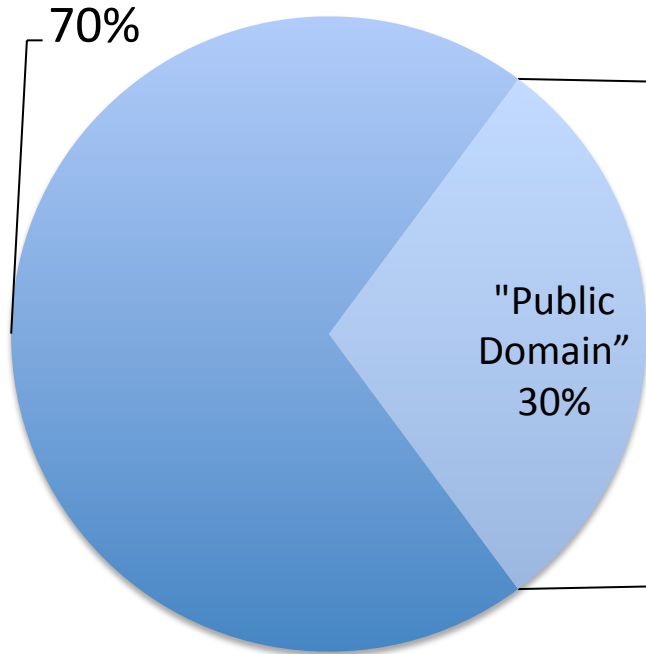
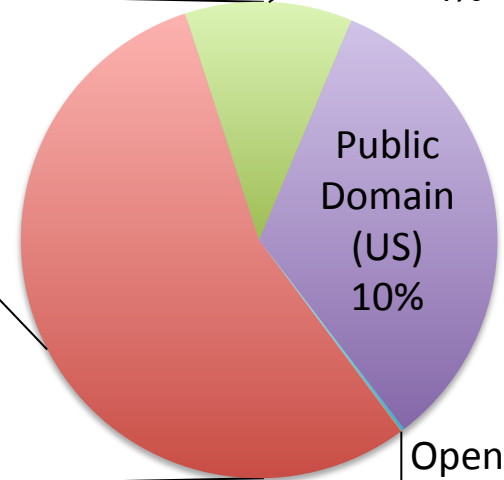# Rights and Agreements

Content Distribution

In-copyright or undetermined
70%

"Public Domain"
30%

U.S. Federal Government Documents (worldwide)
4%

Public Domain (US)
10%

Public Domain (worldwide)
15%

Open Access
.1%

Creative Commons
.01%

Breakdown of HathiTrust book corpus by publication date
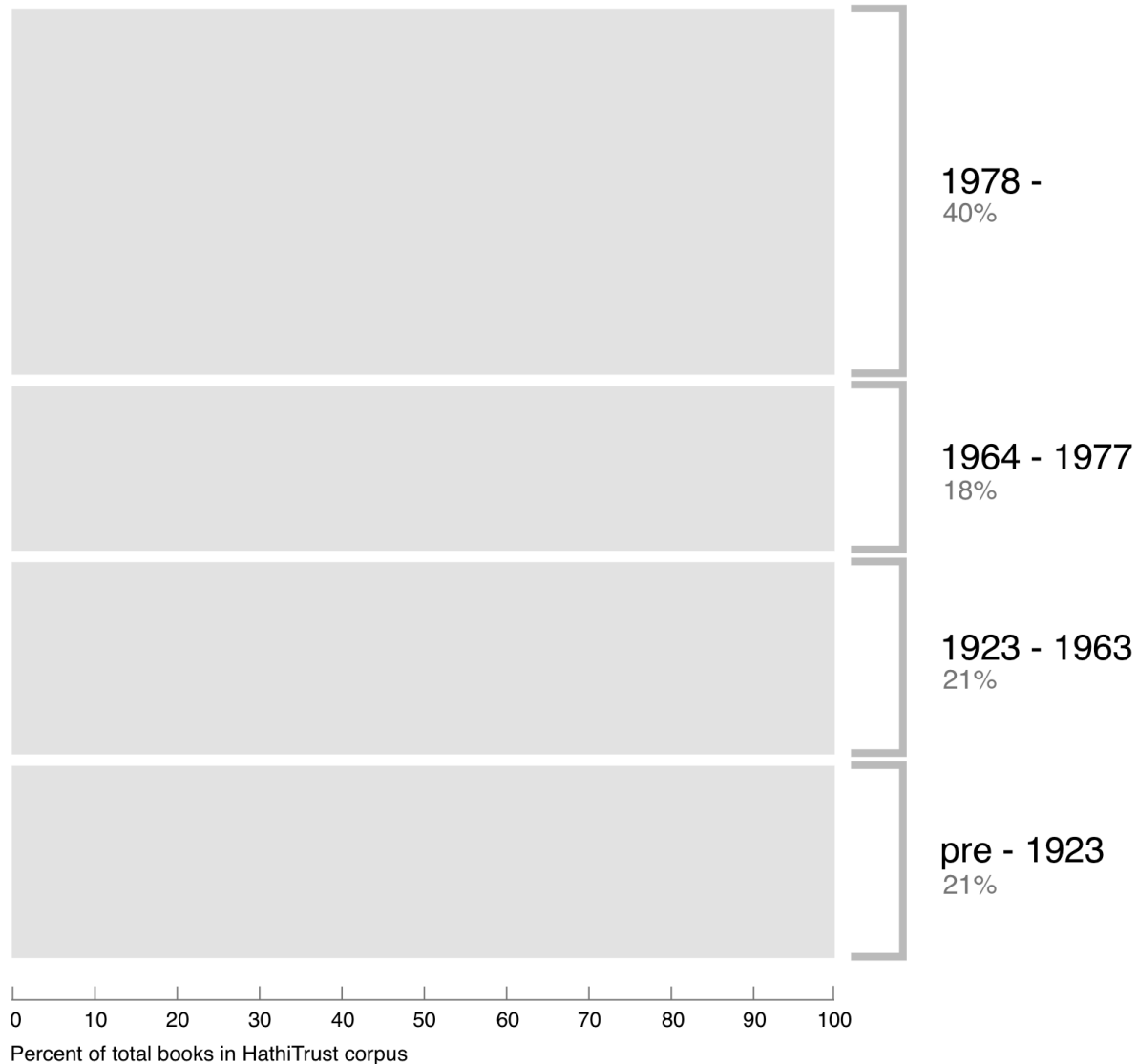
1978 -
40%

1964 - 1977
18%

1923 - 1963
21%

pre - 1923
21%

0 10 20 30 40 50 60 70 80 90 100

Percent of total books in HathiTrust corpus

Breakdown of HathiTrust book corpus by publication date

27%	73%	**1978 -** 40%
28%	72%	**1964 - 1977** 18%
30%	70%	**1923 - 1963** 21%
28%	72%	**pre - 1923** 21%

0 10 20 30 40 50 60 70 80 90 100

Percent of total books in HathiTrust corpus

US Non-US

Copyright status of books published pre-1923 and US works published 1923-1963

Percent of total books in HathiTrust corpus
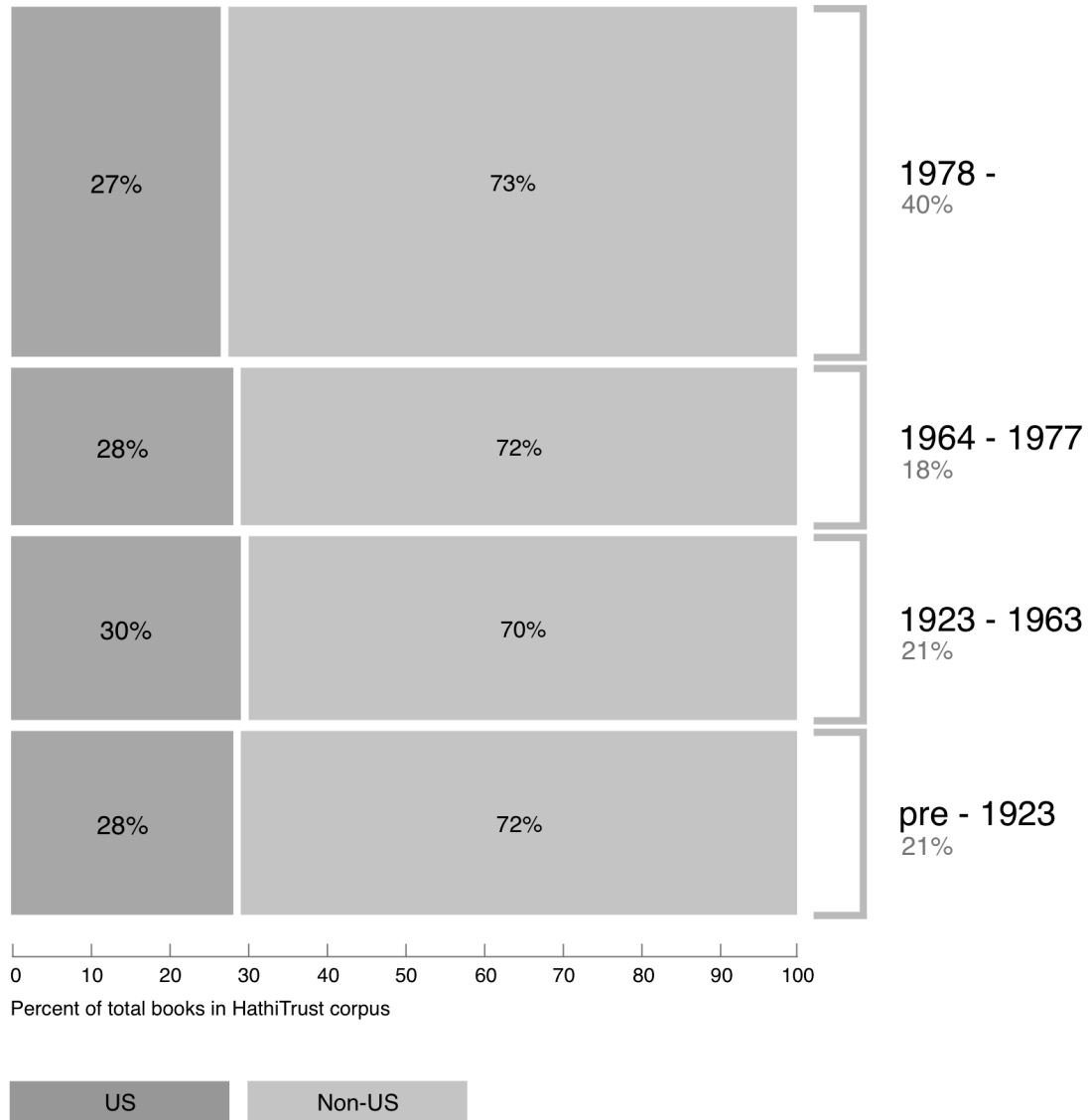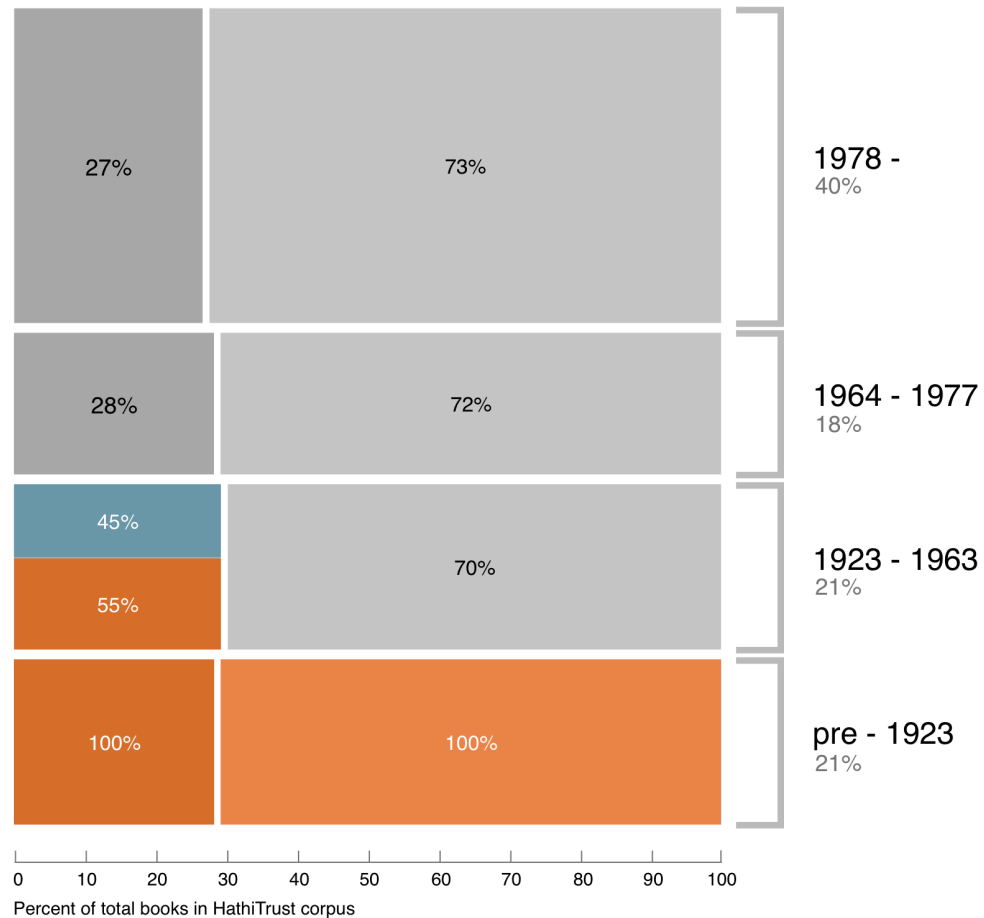
US	Non-US
In Copyright	In Copyright
Orphans	Orphans
Public Domain	Public Domain

Copyright status of books published pre-1923 and US works published 1923-1963

Percent of total books in HathiTrust corpus

	US	Non-US
In Copyright		
Orphans		
Public Domain		

1978 - 40%: US 27%, Non-US 73%
1964 - 1977 18%: US 28%, Non-US 72%
1923 - 1963 21%: In Copyright 45%, Public Domain 55%, Non-US 70%
pre - 1923 21%: Public Domain 100%, Public Domain 100%
Pre-1872 ~ 5%

Copyright status of books published pre-1923 and US works published 1923-1963

Percent of total books in HathiTrust corpus

27%	73%	1978 - 40%
28%	72%	1964 - 1977 18%
45% / 55%	70%	1923 - 1963 21%
100%	100%	pre - 1923 21%

Pre-1872 ~ 5%

Legend:
- US
- Non-US
- In Copyright (US)
- In Copyright (Non-US)
- Orphans (US)
- Orphans (Non-US)
- Public Domain (US)
- Public Domain (Non-US)

Copyright status of books published pre-1923 and US works published 1923-1963

Lawful uses

- Access to users who have print disabilities
- Section 108 uses of materials
- Access to orphan works

Terms of Access

- Available to students, faculty, staff of partnering institutions
  - On library premises or authenticated into HathiTrust
- Partner libraries own a print copy
  - One simultaneous user per print copy owned
- Users must be on U.S. soil
- One page at a time download

Vendor Agreements

- Agreements with vendors common
- Largest impact for HathiTrust is agreement with Google
  - Receive digital copy from Google
  - Share digital copy with partner libraries
  - Prevent download for commercial purposes, redistribution of files, automated or systematic download
- Able to make datasets for research purposes to institutions that sign an agreement with Google

Type of work	Searchable (bibliographic and full-text)	Viewable*	Full-PDF download	Print on Demand	Print disabilities*	Preservation uses (Section 108)*
Public domain worldwide	Worldwide	Worldwide	Partners only if scanned by Google, if not, worldwide.	Worldwide	Partners worldwide	N/A
Public domain (US) – Non-US works published between 1872 and 1923.	Worldwide	When accessed from with the United States	Partners in the US if scanned by Google, if not, anyone US	Available within the United States	Partners in the US; partners worldwide where similar laws in effect	N/A
Works that rights holders have opened access to in HathiTrust	Worldwide	Worldwide	Worldwide (if digitized by Google, full-PDF only available if opened with CC license)	Worldwide with permission	Partners worldwide	N/A
Works that are in-copyright or of undetermined status	Worldwide	Not available	Not available	Not available	Partners in the US; partners worldwide where similar laws in effect	Partners in the US; partner worldwide where similar laws in effect
Orphan works	Worldwide	To participating partners	Not available	Not available	Partners in the US	Partners in the US; partners worldwide where similar laws in effect

* Note: Access to in-copyright works is subject to conditions on Terms of Access slide. See here also.

Key URLs

- Home page – http://www.hathitrust.org
- Data Distribution (including OAI) – http://www.hathitrust.org/data
- Data API – http://www.hathitrust.org/data_api
- Bib API – http://www.hathitrust.org/bib_api
- Hathifiles - http://www.hathitrust.org/hathifiles
- Copyright – http://www.hathitrust.org/copyright
- Access and Use Policies – http://www.hathitrust.org/access_use
- Monthly Updates – http://www.hathitrust.org/updates