



HATHITRUST

A Shared Digital Repository

Big Collections in an Era of Big Copyright: Practical Strategies for Making the Most of Digitized Heritage

Jeremy York
DLF Fall Forum
October 28, 2014



FULL-TEXT

CATALOG

* [Search bar]



LOG IN

Advanced full-text search | Search tips

Full view only

Refine Results

Subject

- [United States](#) (294,339)
- [China](#) (138,451)
- [Agriculture](#) (126,998)
- [Science](#) (117,188)
- [Great Britain](#) (103,775)
- [more...](#)

Author

- [United States. Bureau of the Census.](#) (22,592)
- [Geological Survey \(U.S.\)](#) (17,065)
- [Shakespeare, William, 1564-1616.](#) (13,811)
- [United States. National Aeronautics and Space Administration.](#) (12,053)
- [U.S. Atomic Energy Commission.](#) (11,370)
- [more...](#)

Language

- [English](#) (6,621,288)
- [German](#) (1,306,982)
- [French](#) (1,144,302)
- [Spanish](#) (501,597)

Search Results: 12,542,444 items found for * (Everything)

All Items (12,542,444) Full View (4,664,023)

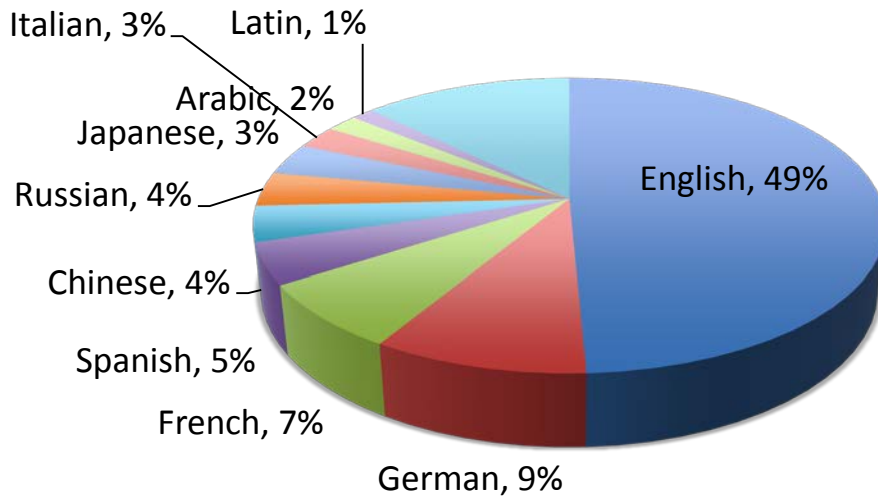
25 per page Prev 1 2 3 4 5 6 7 8 ... 800 Next >>

Select all on page Select Collection Add Selected

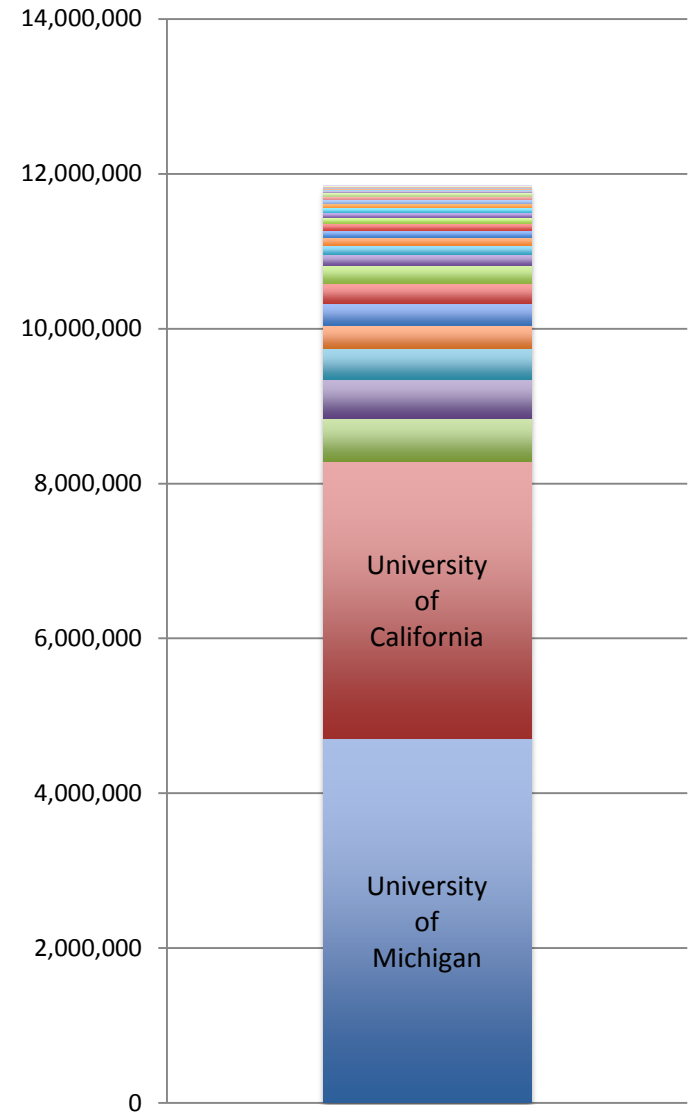
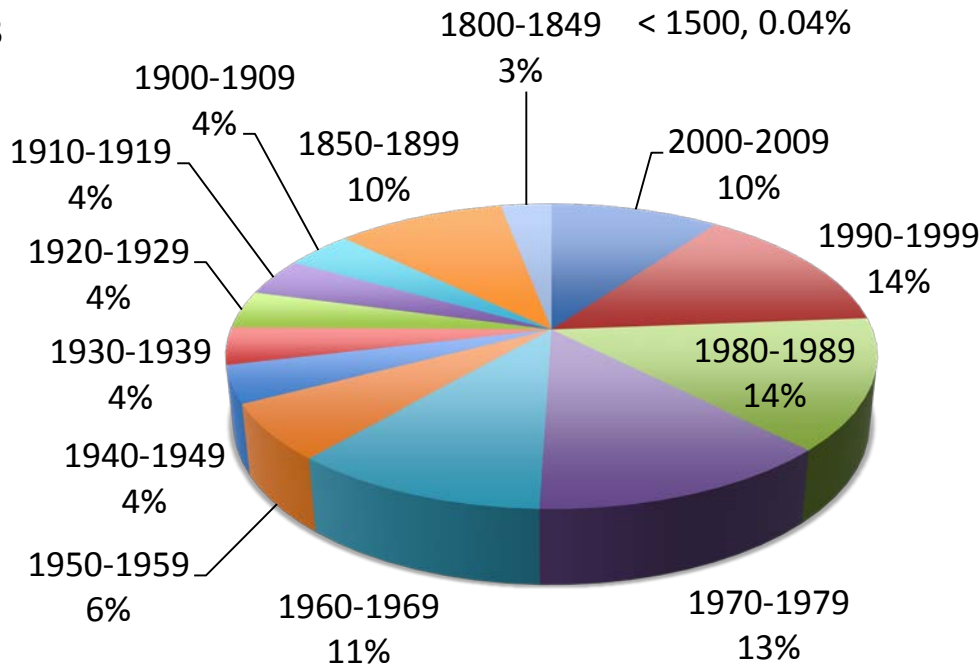
- Gomez Arias : oder, die Mauren der Alpujarren ; ein historischer Roman / von Don Telesforo de Trueba y Cosia ; Aus dem Englischen frei übersetzt von Gustav Sellen. c.1 v.1
by Trueba y Cosío, Joaquín Telesforo de, 1799?-1835.
Published 1829
[Catalog Record](#) [Full view](#)
- Gomez Arias : oder, die Mauren der Alpujarren ; ein historischer Roman / von Don Telesforo de Trueba y Cosia ; Aus dem Englischen frei übersetzt von Gustav Sellen. c.1 v.2
by Trueba y Cosío, Joaquín Telesforo de, 1799?-1835.
Published 1829
[Catalog Record](#) [Full view](#)
- Mosaik : eine Sammlung ernster und launiger Original-Erzählungen. c.1
by Alvensleben, L. von 1800-1868.
Published 1829
[Catalog Record](#) [Full view](#)

12.5 million total volumes
 6.4 million book titles
 327,000 serial titles
 4.6 million volumes in the public domain (~37%)

Top 10 Languages

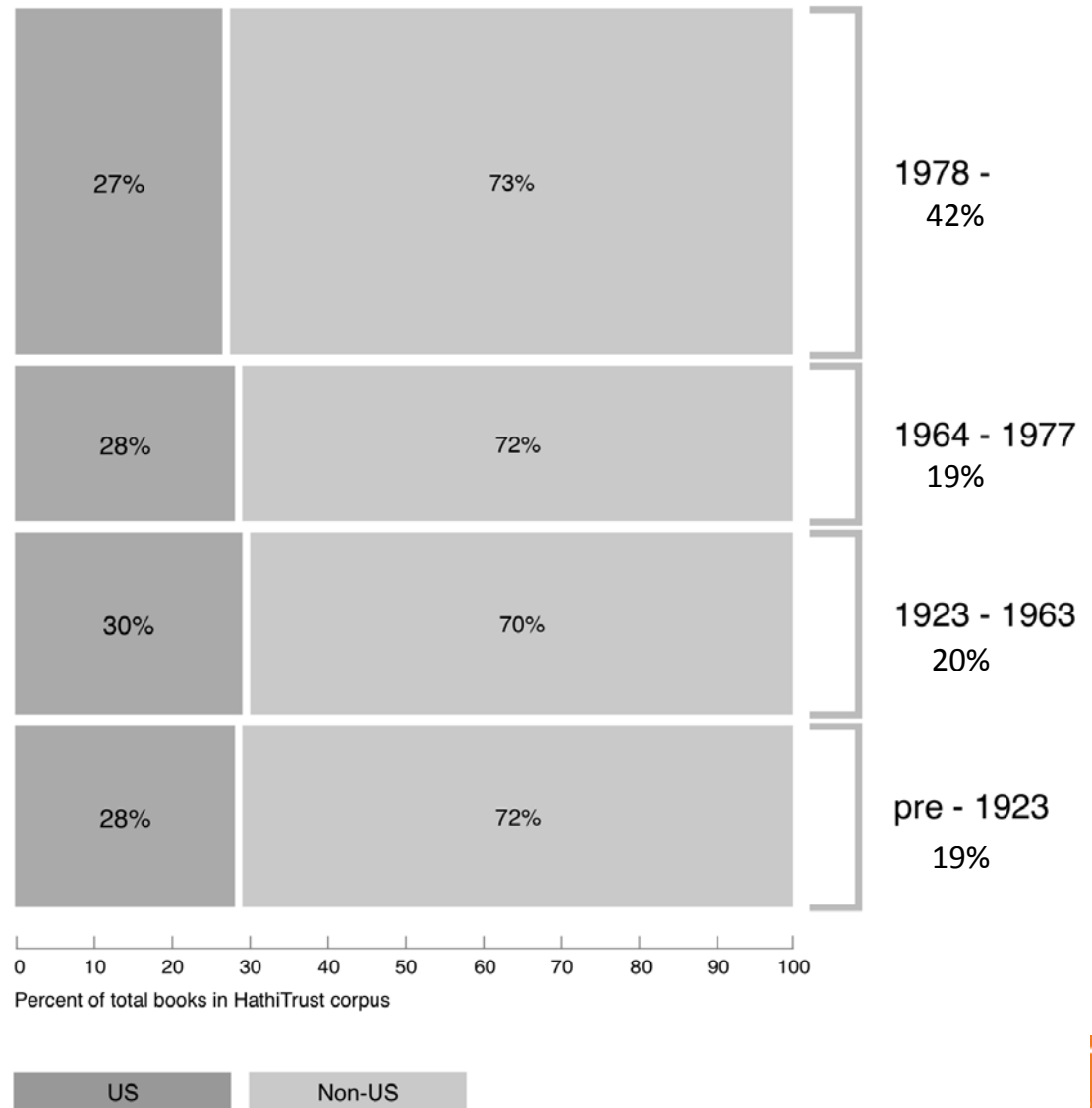


Dates



Breakdown of HathiTrust book corpus by publication date

- Diaries
- Correspondence
- Reports
- Newspapers
- Memoirs
- Books
- Encyclopedias
- Archival materials
- Directories
- Periodicals
- Maps
- Musical scores
- Statistics
- Visual Materials



[Bibliographic Indeterminacy and the Scale of Problems and Opportunities of "Rights" in Digital Collection Building](#) – Wilkin, Feb 2011



The Challenge

- Preserve materials
- Enable the fullest possible use of materials for scholarship and research, and as a public good to the community
 - Bibliographic and Full-text Search
 - Viewing and Download of public domain and open access volumes
 - Collections and APIs
 - Computational Research
 - Print on demand



The Strategy

- Take full advantage of the abilities we have to make collections accessible within the scope of the law
 - Public domain
 - Lawful uses of in-copyright materials



Framework

- The law
- Identification / Copyright determination
- Access policies
- Technical infrastructure



The Law

- If we understand and accept the reasons works should be opened according to the applicable laws, we are willing to open them.



Identification / Copyright Determination

- Automated rights determination

- http://www.hathitrust.org/bib_rights_determination

DateType	008:06
Date1	008:07-10
Date2	008:11-14
PubPlace	008:15-17
PubPlace17	008:17 (last byte of pub place. “u” indicates published in the US, otherwise non-US)
GovPub	008:28
VolDate	Latest year parsed from z30_description field. Set to null if nothing could be parsed or if no z30_description.
BibFmt	Bibliographic record format (BK, SE, etc.)
Imprint field	260 or 264 ind2=1

- Manual rights determinations

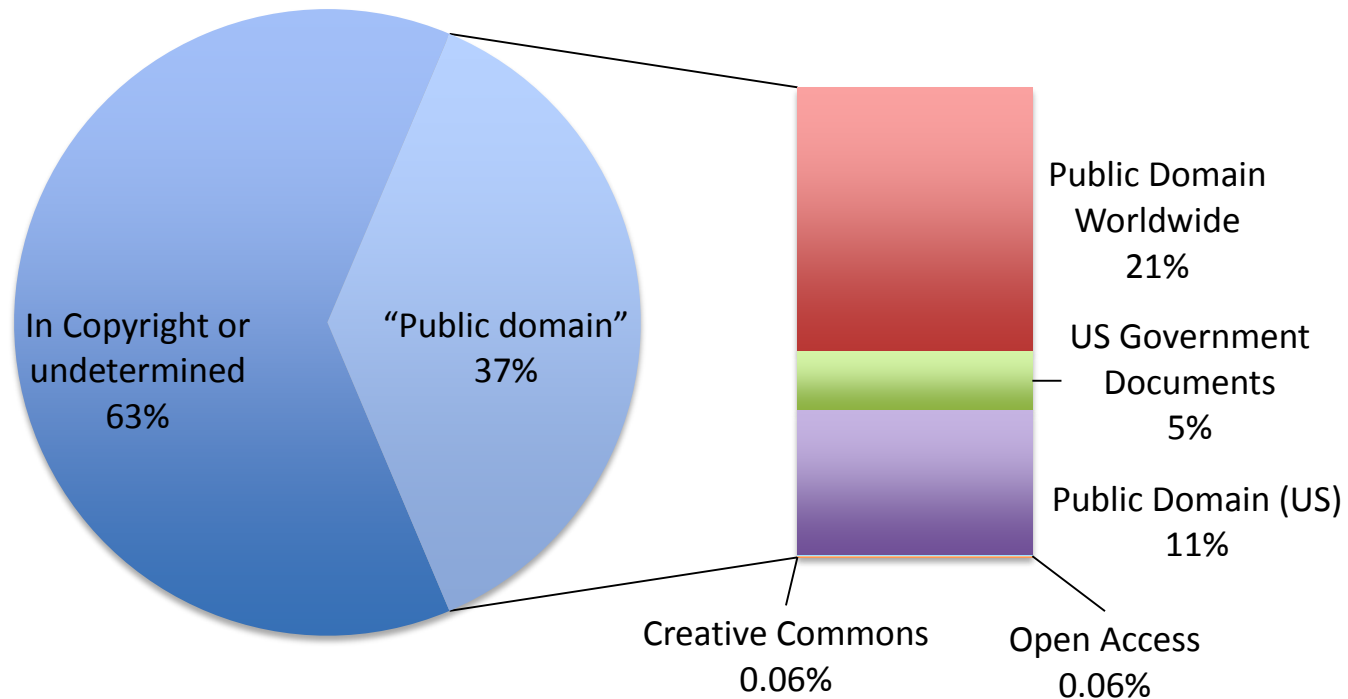


Access Policies

- Public Domain
- Public Domain in the US
- Open Access (including Creative Commons)
- In copyright or Undetermined
- In copyright in the United States
- Nobody (deletions, rights investigations)



Copyright Distribution



Lawful Uses

- Full-text search
- Access for users who have print disabilities:
 - Print copy owned currently or previously; User certified by the partner institution; Accessible to authenticated proxies
- Section 108 (17 USC §108) replacement, preservation, and distribution uses of digital materials:
 - Print copy owned currently or previously; Located within the United States; Replacement copies; access from library premises; Simultaneous accesses determined by print copies held
- Computational Research



Take-downs and Deletions

- Take-down
 - Remove access immediately
 - Investigate rights
 - Re-open or keep closed with new status
- Deletion
 - Rights holder request (contractual obligation)
 - Wholly unusable or superior copy available



Technical Infrastructure

- Strategy for addressing shared problems
- Infrastructure allows/enables
 - Robust discovery
 - Rights determination: automated; distributed manual review
 - Sensitivity to diverse copyright regimes and access policies
 - Storage and management of rights information; availability of information to access systems
 - Rights attributes, and reason codes; system of precedence
 - http://www.hathitrust.org/rights_database
 - Availability of rights information
 - <http://www.hathitrust.org/data>



Framework

- The law ✓
- Identification / Copyright determination ✓
- Access policies ✓
- Technical infrastructure ✓



How to find out more

- 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>
 - RSS http://www.hathitrust.org/updates_rss
- Contact us: feedback@issues.hathitrust.org
- Blogs: <http://www.hathitrust.org/blogs>
 - Large-scale Search
 - Perspectives from HathiTrust

