

3-19-2013

Large Scale Digital Book Initiatives: Checking in with Google Books, HathiTrust, Project Gutenberg and More

Naomi Eichenlaub

Ryerson University, neichenl@ryerson.ca

Follow this and additional works at: http://digitalcommons.ryerson.ca/library_pres



Part of the [Information and Library Science Commons](#)

Recommended Citation

Eichenlaub, Naomi, "Large Scale Digital Book Initiatives: Checking in with Google Books, HathiTrust, Project Gutenberg and More" (2013). *Librarian and Staff Presentations*. Paper 18.

http://digitalcommons.ryerson.ca/library_pres/18

This Conference Presentation is brought to you for free and open access by the Library at Digital Commons @ Ryerson. It has been accepted for inclusion in Librarian and Staff Presentations by an authorized administrator of Digital Commons @ Ryerson. For more information, please contact bcameron@ryerson.ca.

***Large scale digital book initiatives:
Checking in with Google Books,
HathiTrust, Project Gutenberg and more***

Electronic Resources & Libraries
March 19, 2013

Naomi Eichenlaub, Ryerson University, Toronto, ON

Learning objectives

- *Understand the current state of large scale book digitization initiatives such as Google Books, HathiTrust, and more*
- *Determine opportunities and challenges to including this content in your library collections*

Large scale ebook collections

1971: Project Gutenberg

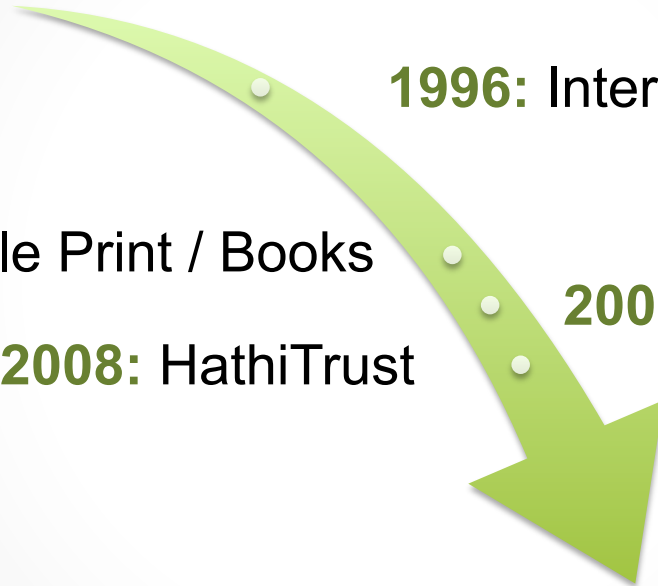
1996: Internet Archive

2004: Google Print / Books

2006: Open Library

2008: HathiTrust

2013: Digital Public Library of America



Google Books



Researching a topic?

Search the latest index of the world's books. Find millions of great books you can preview or read for free.



[Browse books and magazines »](#)

New! Shop for Books on Google Play

Browse the world's largest eBookstore and start reading today on the web, tablet, phone, or ereader.



Google play

[Go to Google Play Now »](#)

[My library](#)

[About Google Books](#) - [Privacy Policy](#) - [Terms of Service](#) - [Blog](#) - [Information for Publishers](#) - [Report an issue](#) - [Help](#) - [Sitemap](#) - [Google Home](#)

©2012 Google

Google Play

The screenshot displays the Google Play website interface. At the top left is the Google Play logo. To its right is a search bar with the text "Search" and a magnifying glass icon. Below the search bar are four navigation tabs: "SHOP", "MY BOOKS", "MY MAGAZINES", "MY MOVIES", and "MY ANDROID APPS".

A large banner features a book cover for "S.E.C.R.E.T." by L. Marie Adeline. The text on the banner reads: "NEW FROM ERONICA ROTH Divergent #3 lands on October 22! Pre-order today" and "S.E.C.R.E.T. by L. Marie Adeline The first in a hot new series". To the right of the book cover, there are smaller book covers for "12 OF NEVER" and "J.P. ROB". Below these covers, the text says "NEW BOOK Pre-order".

Below the banner are two main promotional sections:

- Top Paid:** A list of three books:
 - 1 Indian Horse** by RICHARD WAGAMESE, priced at CA\$13.17 BUY.
 - 2 Safe Haven** by NICHOLAS SPARKS, priced at CA\$8.99 BUY.
 - 3 Fifty Shades of Gre...** by EL JAMES.
- Weekly Deals from \$1.99:** A row of four book covers with their titles and authors:
 - Blind Faith** by CJ LYONS, priced at CA\$8.99 BUY.
 - The Yellow House: ...** by PATRICIA FALVEY, priced at CA\$9.99 BUY.
 - The Wounded Hear...** by ADINA SENFT, priced at CA\$2.99 BUY.
 - So Cold the River** by MICHAEL KORYTA, priced at CA\$6.99 BUY.

"[T]he massive scale of library digitization enterprises, along with the high costs of digital preservation, demand a web-scale collaborative solution for ensuring long-term access to the digital output and a new vision for a collective collection." (Christenson, 2011)

HathiTrust



HATHI
TRUST Digital Library

[Help](#) [Feedback](#)

[Home](#) [About](#) [Collections](#) [My Collections](#)

① Catalog Search

Search information **about** the items.

Full view only

[Advanced Catalog Search](#) | [Search Tips](#)

📖 Full-text Search

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

Full view only

[Advanced Full-text Search](#) | [Search Tips](#)

📁 Collections

Browse, search, or make HathiTrust 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

[Try our mobile website!](#)

[Information about the Authors Guild Lawsuit](#)

- [Putting It All Together: HathiTrust Vision, Practice, and Implementation](#)
- [HathiTrust: An Author Never Forgets](#)
- [Update on January 2013 Activities](#)
- [Towards Practical Relevance Ranking for 10](#)

Currently Digitized

10,645,320 total volumes
5,599,403 book titles
277,270 serial titles
3,725,862,000 pages
477 terabytes
126 miles
8,649 tons
3,307,124 volumes (~31% of total) in the public domain
[View visualizations of HathiTrust](#)

Featured Collection



UM Press

The Univ. of Michigan Press titles available in HathiTrust

HathiTrust

In copyright or undetermined	70%
Public domain or open access	30%
Public domain worldwide	19%*
Public domain in the US	10%
Open Access	1% **

* Of which approximately 4% are US federal gov't documents

* * Approximately 10,000 volumes (including Creative Commons-licensed)

<http://www.hathitrust.org/documents/HathiTrust-Overview-Handout.pdf>

http://www.hathitrust.org/access_use

HathiTrust: Visualizations

Call Numbers - Public Domain

http://www.hathitrust.org/visualizations_callnumbers_pd

Languages - Public Domain

http://www.hathitrust.org/visualizations_languages_pd

Dates - Public Domain

http://www.hathitrust.org/visualizations_dates_pd

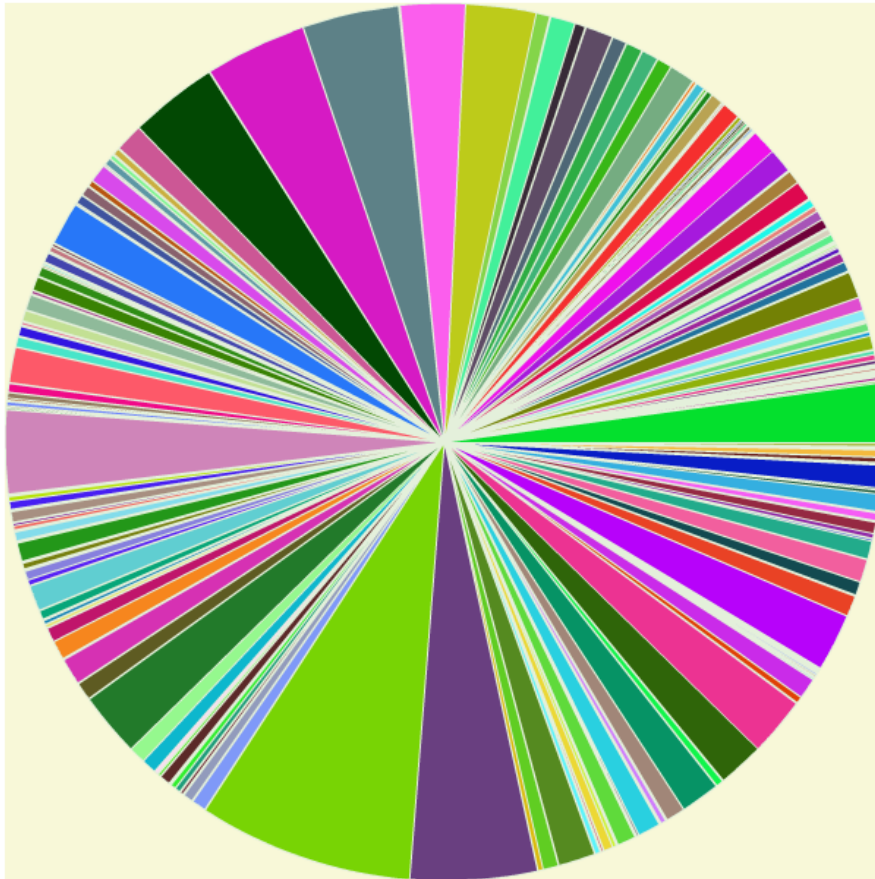
http://www.hathitrust.org/statistics_visualizations

HathiTrust: Visualizations

[About](#) > [Our Digital Library](#) > [Statistics and Visualizations](#) > [Call Numbers - public domain](#)

HathiTrust Call Numbers - Public Domain

A visualization of HathiTrust **public domain** works by call number (Library of Congress classification).
[View statistics information >>](#)



LC Classification	Count	Percent
A - General Works	95	0.01
AC - Collections. Series.	1,444	0.19
AE - Encyclopedias	248	0.03
AG - Dictionaries and other reference	234	0.03
AI - Indexes	67	0.01
AM - Museums. Collectors and collecting	192	0.03
AN - Newspapers	4	0.00
AP - Periodicals	2,122	0.28
AS - Academies and learned societies	1,382	0.18
AY - Yearbooks. Almanacs. Directories	450	0.06
AZ - History of scholarship and learning	77	0.01
B - Philosophy; Psychology; Religion	6,591	0.87
BC - Logic	528	0.07
BD - Speculative philosophy	1,175	0.16
BF - Psychology	4,863	0.65
BH - Aesthetics	253	0.03
BJ - Ethics	2,263	0.30
BL - Religions. Mythology. Rationalism	3,407	0.45
BM - Judaism	1,451	0.19
BP - Islam. Bahaim. Theosophy	671	0.09
BQ - Buddhism	104	0.01
BR - Christianity	5,056	0.67
BS - The Bible	6,476	0.86

Challenges

Copyright /
lawsuits

Orphan
works

Metadata

Language

Quality

Search

Open access ebook initiatives

- OAPEN Library

<http://www.oapen.org/>

- DOAB

<http://www.doabooks.org/>

- Unglue.it

<https://unglue.it/>

Collection Development

How can you get this ebook content into your library collections?

Conclusions

"The large-scale digitization of books [...] is generating extraordinary collections of intellectual content that are transforming teaching and scholarship at all levels of the educational enterprise". (Conway, 2010)

<http://books.google.com/ngrams/>

Works cited

Christenson, Heather. 2011. "HathiTrust: A research library at web scale". *Library Resources & Technical Services*. 55:2.

Conway, Paul. 2010. "Measuring content quality in a preservation repository: HathiTrust and large-scale book digitization". *Proceedings of 7th International Conference on Preservation of Digital Objects, iPres 2010*, 19-24 Sept. 2010, Vienna, Austria, pp. 95-102 <http://hdl.handle.net/2027.42/85227>

Gooding, Paul. 2012. "Mass digitization and the garbage dump: The conflicting needs of quantitative and qualitative methods." *Literary and Linguistic Computing* 27:4.

New, William. 02/08/2013. "Fate of Google Book Search still to come, expert says". *Intellectual Property Watch*.

Nunberg, Geoff. 2009. "Google Books: A metadata train wreck". *Language log*. August 29, 2009: <http://languagelog.lidc.upenn.edu/nll/?p=1701>

Polanka, Sue. 2013. "Open access ebooks". *Online*. 37:1.

Polanka, Sue. 2012. "Ungluing open access ebooks". *Online*. 36:3.

Thank you!



Naomi Eichenlaub
Ryerson University (Toronto, ON)
neichenl@ryerson.ca