14 Million Books & 6 Million Visitors:

HathiTrust Growth and Usage in 2016

The Growth of the HathiTrust Collection in 2016

The HathiTrust collection continues to grow steadily. As of January 1st, 2017, there are 14,816,187 volumes in the collection. Over one million volumes were added to the collection over the course of the preceding year, scanned from the library collections of 39 contributors. These included several new unique collections, such as:

- [Latin American cultural and historical materials from the Nettie Lee Benson Latin American Collection at the University of Texas at Austin](#) (See [the press release](#) for more information.)
- [Hebrew Manuscripts from the University of Maryland](#)
- [Materials related to Baptist organizations and history, from Wake Forest University](#)
- [Books digitized from the general collection at Michigan State University](#)
- [Books related to the history of Catholicism from the University of Notre Dame](#)
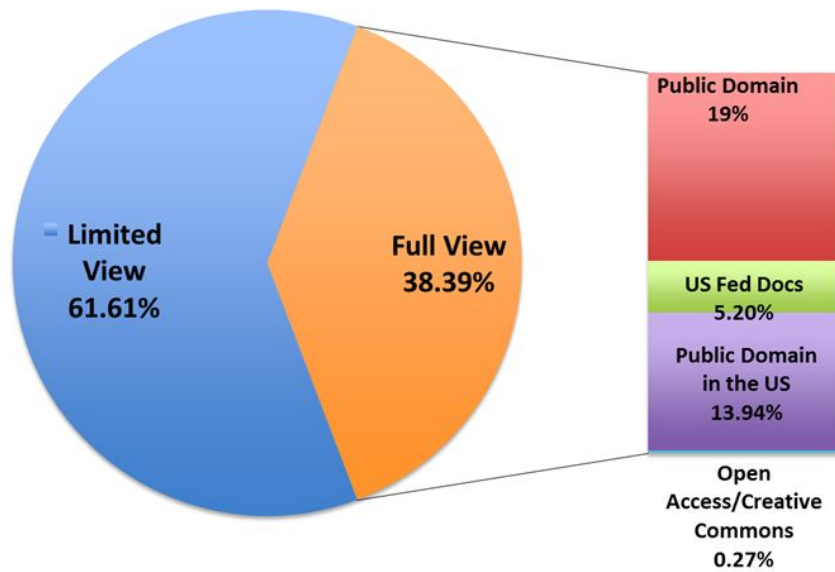- [University of Queensland publications](#)

Books made fully viewable in HathiTrust in 2016

HathiTrust makes books available through a variety of methods. New content added to the collection may be identified as falling within the public domain, and those items are made available in full view to users immediately upon ingest. Copyright review projects (through the CRMS program) continued throughout 2016, which opened another 29,000 items. We received approximately 75 permissions agreements throughout 2016, giving us permission to open over 15,000 volumes (12,700 of these were opened with Creative Commons licenses). In addition, contributors added 128,808 United States federal documents to the HathiTrust collection, the majority of which are in the public domain. We ended the year with 38% of the collection available to users (5.7 million volumes).

The remaining 62% of the collection is preserved and made available for other purposes: all users can search across and within those limited-view books, and qualified users with print disabilities can access the full text. In addition, in 2016 the [HathiTrust Research Center began providing non-consumptive research services for text and mining across the entire HathiTrust collection, including copyrighted titles](#).

Breakdown of Rights and Access in HathiTrust

[Caption: The pie chart above describes two main categories of materials in the HathiTrust collection: limited-view materials that make up 61.61% of the collection, and full-view materials that make up 38.39% of the collection. The bar chart breaks down the full-view materials into further subsets: public domain materials are 19% of the collection; United States federal documents are 5.2%; books that are only in the public domain within the United States account for 13.94%; and books that have been opened with permissions agreements account for 0.27%.]

Usage of the HathiTrust Collection

From a review at our usage data, we can offer some high-level information about who our users are and what their habits are. Over 6.17 million users visited our site over the course of 2016, culminating in 10.92 million sessions.

Who are our users?

About 49% of our users were located in the United States in 2016. This was a rise from 46% in 2015, but the percentage has stayed at approximately 50% for years. The remaining 51% of users are scattered across a long tail of 236 other nations, topped by the United Kingdom, Canada, Germany, Italy, India, France, China, Australia, and Spain.

Our users are primarily English speakers, as detected by the language of their browsers. 64% read American English, and 6% prefer British English. The remaining top 10 languages include: two variants of German, Spanish, French, Chinese, two different variants of Italian, and Russian.

That the largest percentage of users are from English-speaking countries is not surprising. Although the HathiTrust collection contains a vast amount of material in non-English languages, the website interface is entirely in English, and the catalog records for most of the materials were created by English-speaking catalogers in the United States and Canada. (There are two notable exceptions to this: materials from Keio University in Japan and Universidad Complutense de Madrid in Spain.) Also unsurprising: users from countries where the primary language is not English have a higher bounce rate (i.e., how frequently users leave a website without exploring beyond their entry page), typically higher by 10 percentage points or more.

Country	% of Total Sessions	Bounce Rate
United States	49.54%	20.59%
United Kingdom	6.00%	24.35%
Canada	3.65%	22.66%
Germany	3.43%	27.53%
Italy	2.72%	36.97%
India	2.14%	42.02%
France	2.01%	37.95%
China	1.99%	59.18%
Australia	1.93%	21.49%
Spain	1.68%	32.27%

Users logged in to access personalized services during sessions about 3.4% of the time, and 2.9% (or 186,969) of all sessions are affiliated with HathiTrust members. In the remaining 96% of sessions, users did not log in. Logging in provides users with the ability to build collections and, for those users affiliated with member institutions, the ability to download full pdfs where there are download restrictions.

This is a fascinating statistic, but we need to investigate user behavior more closely to better understand it. Most of our users are in the general public and not affiliated with member institutions. It may be that those users are so numerous that they simply dwarf usage from members. But are users at our member institutions satisfied with reading on-screen only? Are they able to get what they need from reading only a few screens? Or does our user interface not clearly guide them to services available only with authentication or provide sufficient signals that their university is a HathiTrust member? We will need to dive into our usage data in future analyses in order to understand this behavior.

How do our users find us?

Our users find us through a variety of means. Most come to us either through search engines (40%) or through direct referrals (39%). By far, our largest source of referral traffic is the Online Books Page, an indexed listing of books that are freely available on the Internet. The site is hosted at the University of Pennsylvania and managed by John Mark Ockerbloom. Clicks from this source account for a whopping 23% of our referral traffic, or 997,000+ sessions.

Other top sources of referral traffic include (in no particular order): WorldCat, Wikipedia, the Digital Public Library of America (where we have contributed records for all of our open access content), the National Library of Australia, and the online catalogs for a number of libraries: the University of Michigan, the Karlsruher Institut für Technologie, the University of Hong Kong, and Harvard University.

A significant portion of users (20%) find their way directly to us, either through saved bookmarks or links shared in ways that are difficult for Google Analytics to track (e.g., a user clicks a link embedded in a PDF document).

Finally, users are constantly sharing links to HathiTrust contents on many social media platforms. These referrals only account for 1.16% of our referral traffic but can provide fascinating glimpses into how users are using HathiTrust content in non-academic settings. The top social media sites where users are sharing HathiTrust content are: Facebook, Blogger, Twitter, reddit, Pinterest, VKontakte, WordPress, LibraryThing, YouTube, and Sina Weibo.

- Blogging platforms showcase the work of non-academic historians who [use sources in HathiTrust to research local history](#); writers who [use HathiTrust to research historical periods within which their characters live](#); teachers who [share links to resources on specific subjects](#), students who [create digital scholarship websites to document their research projects](#), and so many more cases.

- Users on Reddit engaged in [long, discursive conversations that included references to a speech Robert Heinlein gave before Congress](#) in the context of a larger discussion about "why are we trying to colonize Mars?"
- Crafters on [Pinterest](#) and [Ravelry](#) share links to old crochet, knitting, embroidery, etc., patterns.
- Genealogists [share links to resources they have discovered](#).
- And of course, users everywhere just want to [highlight books they have discovered](#) or [include in their carefully curated bibliographies on very specific subjects](#).

What are our users reading?

Once users have arrived at our site, what are they reading? Following are our top titles that people view. Some of these top titles are curious phenomena, such as in the case of *The Surnames of Scotland*, which became an extremely popular book this year after it was cited on the genealogy website, ScotlandsPeople. Curiously enough, most of this traffic was generated within the 24 hours after ScotlandsPeople first posted about this book - and yet, such a large number of visitors clicked that link, that this book made it to the very top of our most popular titles for the entire year.

Quicksand, a novel by Nella Larsen that tells the story of a mixed race woman who searches for a home in many different communities and locations, has consistently been at the top of our top 10 lists for several years now. *The Human Figure*, by John H. Vanderpoel is also another frequently visited book. It is a book instructing artists how to draw and paint the human figure.

Top 10 Books in 2016
[The surnames of Scotland, their origin meaning and history, by George F. Black](#)
[Quicksand, by Nella Larsen](#)
[The Marines Magazine. v. 2 (1916-17)](#)
[The Ordeal of Mansart, by W. E. B. Du Bois](#)
[Families Directly Descended From All the Royal Families in Europe (495 to 1932) and Mayflower Descendants, By Their Lineal Descendant, by Mrs. (Oscar Herbert) Elizabeth M. Leach Rixford](#)
[Il Poligrafo Domenica. v. 3 (Jan.-June 1812)](#)

[M. Tullii Ciceronis Ad M. Brutum orator. Eine kritische und erklarende schulausgabe von Carl Peter und Gottob Weller](#)

[Benjamin Chew, 1722-1810, Head of the Pennsylvania Judiciary System Under Colony and Commonwealth, by Burton Konkle](#)

[The Human Figure, by John H. Vanderpoel](#)

[Soviet-American Relations: the Détente Years, 1969-1972](#)

Conclusion

From this quick overview of usage, we see that HathiTrust users come from all over, have many different needs, and many differents usage habits. At such a high level, it is difficult to assess on how well the HathiTrust Digital Library is helping users meet their needs. Future analysis may benefit from breaking users into broad categories to see how they interact with the library, which may feed into our development of new features and services and improvement of existing features and services.

It seems clear from the number of visitors that, in its 8th year of existence, HathiTrust serves a crucial need for users around the world. We plan to continue to improve our ability to meet their needs, through the collection of additional materials (in particular rarely held and special collections materials, which are now the focus of digitization projects in many libraries), through development of better methods of tracking on and improving the quality of digital content, and through more robust methods for providing access to users with print disabilities. In order to receive alerts about development and improvement of our services, please subscribe to our newsletter at http://eepurl.com/cxjNWT.