

The background features a large, light-colored watermark of the HathiTrust logo, which includes a stylized elephant head and the text "HATHI TRUST" and "Digital library".

HathiTrust- History, Content & Tools for Research: in the case of CJK materials

2017 PRRLA Conference

Todd Grappone grappone@library.ucla.edu

Su Chen suchen11@library.ucla.edu

Hangzhou, China

October 17, 2017

What is HathiTrust?

- Brief history: 2008-
 - Partners
 - Five consortia (Big Ten Academic Alliance, Claremont University Consortium, State University System of Florida, University of California, U of Texas system)
 - Individual partners: 126
 - International partners
 - Keio University (Japan), Japanese and Chinese
 - Universidad Complutense de Madrid (Spain), Spanish
 - University of Queensland (Australia), English
 - University of Alberta (Canada), English, French
 - Google digitized resources plus others
- Membership fees

HathiTrust



www.hathitrust.org

Contents in the Hathitrust

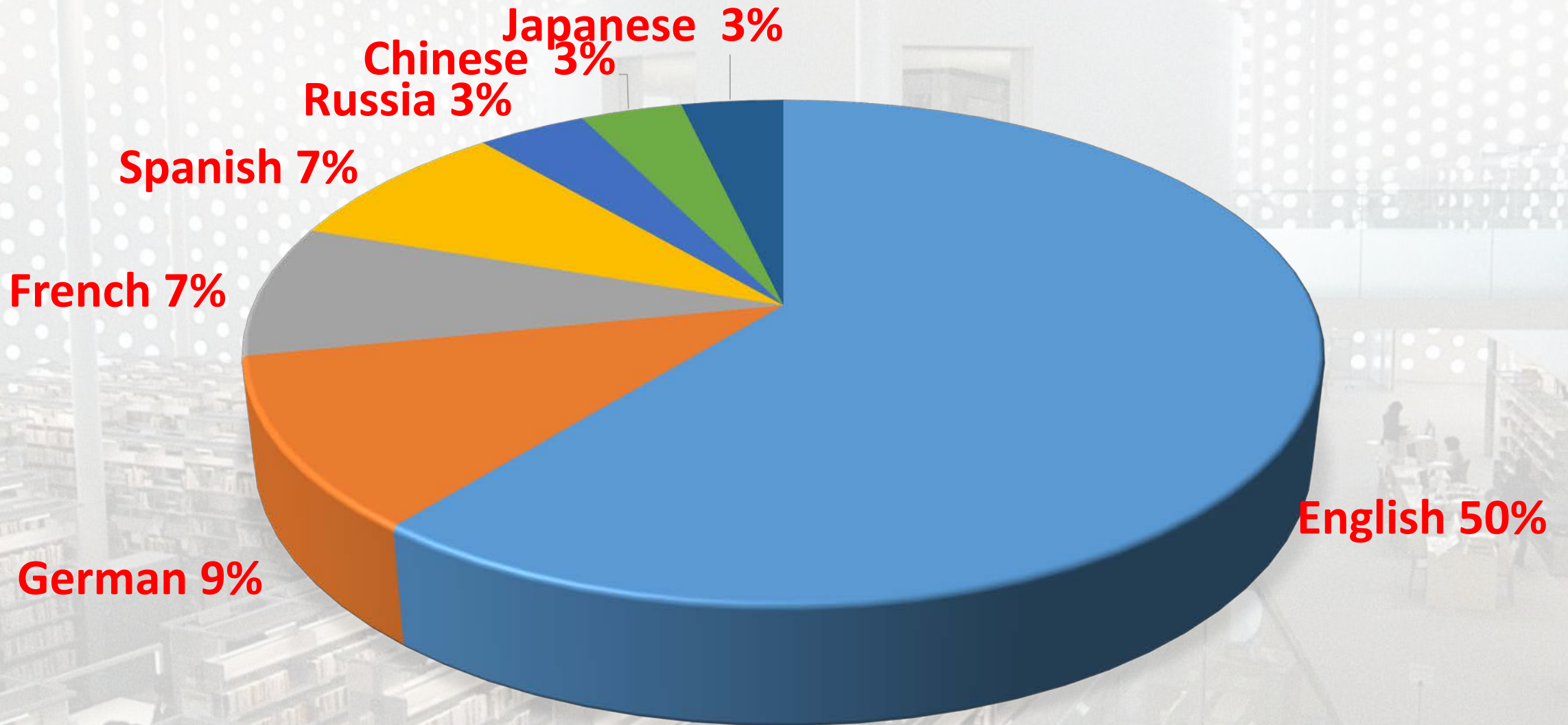
- 8 million titles, 16 million volumes
 - Both limited view and full-view
- Original format
 - Book: 7 million titles
 - Others: 1 million
- Major content contributors
 - U of Michigan: near 3 million
 - UC system: over 2 million
 - Harvard University: 504,000

14 Million Books & 6 Million Visitors in 2016

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 2.9% (or 186,969) of all sessions are affiliated with HathiTrust members. In the re of sessions, users did not log in. Logging in provides users with the ability to bu and, for those users affiliated with member institutions, the ability to download fu there are download restrictions.

Top language materials in Hathitrust



What about CJK materials in HathiTrust?

- Are these new developments applied to Chinese Japanese & Korean materials?
- Currently, HathiTrust holds,
 - Chinese 263,000 titles, 547,000 items
 - Japanese 235,000 titles, 510,000 items
 - Korean 31,000 titles, 57,000 items
- The CJK in HathiTrust is larger than UCLA EAL holdings, close to UCB's holding, what is the contributions to the East Asian studies?

How the CJK materials are used?

- None of the major East Asian digital humanities projects at Chicago, Columbia, Cornell, Harvard, Indiana, Michigan, Princeton, Stanford, Virginia, UC Berkeley, UCLA and Yale have used it.
- Why?
 - Not known widely among librarians
 - Not know widely among scholars
 - Undiscoverable in the UCLA Catalog
 - No easy way to dedup
 - Scan quality is limited
 - Metadata is limited
 - No OCR, unable to accommodate keyword search
 - Unable to apply any DH's tools
 - Some pre-1923 books can be viewed, downloaded and printed when no holdings found at UCLA
 - Some newly published books can be found when authors give up copyright
- Summary
 - The CJK materials in Hathitrust has limited value for East Asian studies if the above issues are not addressed.

THANK YOU!
Questions?



국립중앙도서관

디지털도서관
THE NATIONAL DIGITAL LIBRARY OF KOREA