Overview of today’s webinar

- UC’s mass digitization projects
- What is the HathiTrust and why is UC participating?
- What’s in the HathiTrust?
- A brief tour of current HathiTrust services
- How to find out more, & questions
UC mass digitization efforts

- October 2005 to present
- 2 digitization partners, multiple sources of funding
- 10 campuses* + CDL
- 7 different project locations – NRLF, SRLF, UCB, UCLA, UCSC, UCSD, UCD
- CDL in coordinating role
- Over 2.6 M volumes digitized

UC’s mass digitized content

- Open Content Alliance / Microsoft / Internet Archive ~ 200K vols
  - Out-of-copyright works
  - Primarily English language, some romance languages
  - Library and grant-funded
- Google ~ 2.5 M+ vols
  - In-copyright and out-of-copyright works
  - All languages
  - Almost entirely funded by Google
Where are our digital books?

HathiTrust
- Copies of approximately 1.1 million UC Google books are stored on HathiTrust servers at the University of Michigan and Indiana University, and an additional copy is backed up on tape. More are being added daily
- Internet Archive books are being added to HathiTrust

Internet Archive
- Internet Archive keeps copies of the books they digitize

Google
- Google keeps copies of the books they digitize

Where can you find UC books?

- UC-eLinks
- Open Library: [http://openlibrary.org/](http://openlibrary.org/)
RECHERCHES

DES ANCIENNES

CORPORATIONS

Ouvrières et Marchandes

DE LA VILLE DE REMIRE

PAR

ARMAND BERULLON

LIBRÉ BRUXELLES
What is the HathiTrust?

A shared digital repository for mass digitized content, founded in October 2008
- A partnership of major U.S. research libraries
- Operates at web scale
- Currently stored as of April 19, 2010
  - 5,643,119 volumes
  - 1,975,091,650 pages
  - 210 terabytes
  - 67 miles
  - 4,585 tons
  - 864,747 volumes (~15% of total) in the public domain
- Over 1.1 million UC volumes, and counting...

The University of California Libraries
Why is UC participating in the HathiTrust?

- Preservation and stewardship of UC resources
  - Brings our Google and Internet Archive books together in a common preservation repository under UC control
- Economy of scale
  - Storing mass digitized books is expensive – many terabytes of data
- Better access to our own books
  - Create robust links to full text in HathiTrust in Next Generation Melvyl, including all viewable content from UC and other participating libraries
  - Build improved access interfaces via the HathiTrust API
- Aggregate multiple library collections for greater research impact
  - HathiTrust will support shared access and search mechanisms across all partner content to the extent possible
  - With UC, over 5 million books and counting
  - Over ¾ million books in the public domain
- Experiment with large scale search, text mining, and other specialized services developed with academic users in mind
  - Google and Internet Archive are building services for the general user
  - Research libraries will build services optimized for serious research

HathiTrust Partners

- Lead partners
  - University of Michigan
  - University of California
  - Indiana University

- Members
  - Columbia University
  - Indiana University
  - Michigan State University
  - Northwestern University
  - The Ohio State University
  - Penn State University
  - Purdue University
  - University of California Berkeley
  - University of California Davis
  - University of California Irvine
  - University of California Los Angeles
  - University of California Merced
  - University of California Riverside
  - University of California San Diego
  - University of California San Francisco
  - University of California Santa Barbara
  - University of California Santa Cruz
  - California Digital Library
  - The University of Chicago
  - University of Illinois
  - University of Illinois at Chicago
  - The University of Iowa
  - University of Michigan
  - University of Minnesota
  - University of Wisconsin-Madison
  - University of Virginia
UC involved at many levels

Executive Committee
- Brian Schottlaender, UCSD
- Laine Farley, CDL

Strategic Advisory Board
- Bernie Hurley, UCB
- R. Bruce Miller, UCM
- Patricia Cruse, CDL

Working Groups
- CDL & campus participants according to tasks

Operations & Development
- CDL, for now
- Potentially campuses?

What’s in the HathiTrust?

- Volumes from major research libraries, digitized by Google, Internet Archive, and other agents
- 5,643,119 preserved volumes
- 864,747 volumes (~15% of total) in the public domain & viewable

- Books
- Journals
- Serials
- Government documents
- Other “book like” items
Context for HathiTrust services

- Backend based on original work by the University of Michigan
- CDL participating in development
- HathiTrust Discovery Interface working group engaged in development of “HathiTrust WorldCat Local”
- Development priorities going forward will be collaboratively determined
- HathiTrust offers data to build local services
- Very early days of user experience…
The University of California Libraries

HathiTrust Digital Library

Catalog Search

Search information about the item.

Advanced Catalog Search

Search Tips:

Phrase Searching: Use quotes to search for exact phrases, e.g., "small flour".

Wildcards: Use "*" or "%" to search for alternate forms of a word. Use "*" to replace multiple characters, and "%" for a single character, e.g., "sp*nt" will find "spent", "spant", etc., but not "spent", etc., and "%" will find "onto", "tople", etc.

Boolean Searching: Use AND and OR between words to combine them with Boolean logic, e.g., "heart AND surgery" and "surgery" will find items about heart surgery or coronary surgery.

Multiple Term Searching: When you search for phrases, wildcards, or common words (such as "and", "or", "the", etc.) to expand your search.

Note: Printed search does not currently include functionality in use elsewhere in HathiTrust (e.g., there is no sorting or collection related functionality).

Full-text Search

Search words that occur within the items.

Collections

Browse, search, or make HathiTrust collections.

Featured Collection:

Ancestry and Genealogy

Online content related to genealogy, family history, ancestry, etc.

Advanced Catalog Search:

AND

Limit To:

Format:

Video

Audio

Text

Image

Audio Video

Audio Text

Audio Image

Text Image

Video Image

Audio Video Image
<table>
<thead>
<tr>
<th>Date of Publication</th>
<th>The anatomy of knowledge; papers presented to the Study Group on Foundations of Cultural Unity, Bowdoin College, 1965 and 1966. Edited by Marjorie Gran. Published 1969. (Limited</th>
<th>Search Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date of Publication</td>
<td>Texte des communications. Berichte, Communications, by Word congress of universal documentation, Paris, 1937. Published 1937. (Limited</td>
<td>Search Results</td>
</tr>
<tr>
<td>Date of Publication</td>
<td>The history and philosophy of knowledge of the brain and its functions; an Anglo-American symposium, London, July 15th-17th, 1967. Sponsored by the Wellesley Historical Medical... by Symposium on the History and Philosophy of Knowledge of the Brain and Its Functions, London, Published 1968.</td>
<td>Search Results</td>
</tr>
<tr>
<td>Date of Publication</td>
<td>Classification, Universal and old world history, by Library Congress. Classification Division. Published 1921.</td>
<td>Search Results</td>
</tr>
<tr>
<td>Date of Publication</td>
<td>Proceedings of the ... International Conference on Information and Knowledge Management... by International Conference on Information and Knowledge Management. Published 1999. Multiple volumes.</td>
<td>Search Results</td>
</tr>
<tr>
<td>Location</td>
<td>University of California (1966)</td>
<td>Search Results</td>
</tr>
</tbody>
</table>
LIBRARY
OF
UNIVERSAL KNOWLEDGE.
A REPRINT
OF THE LAST (1896) EDINBURGH AND LONDON EDITION
OF CHAMBERS'S ENCYCLOPEDIA,
it: Cephas Woollens & "W"t"nms (Editors),
FIFTEEN VOLUMES,
VOLUME IV,
NEW YORK,
AMERICAN BOOK EXCHANGE,
TRIBECE BUILDING, 1890.
Library of universal knowledge. A reprint of the

A REPRINT
OF THE LAST (1880) EDINBURGH AND LONDON EDITION
OF CHAMBER'S ENCYCLOPEDIA.
For more information

- HathiTrust web site: http://www.hathitrust.org/
- HathiTrust monthly updates: http://www.hathitrust.org/updates
- CDLInfo: http://www.cdlib.org/cdlinfo/
- CDL web site: http://www.cdlib.org/services/collections/massdig/
- http://www.cdlib.org/services/hathi/index.html
- Ask Heather! heather.christenson@ucop.edu

Questions??

The University of California Libraries