HathiTrust discovery and access @ Harvard Library
Corinna Baksik
October 26, 2020
Code of Conduct

HathiTrust events provide an inclusive environment that welcomes inquiry, constructive criticism and debate, and candor. HathiTrust does not tolerate personal attacks, harassment of any kind, verbal or physical violence, or disruptive behavior. All attendees are expected to be respectful of our community’s diversity and generous of others’ views. A full Code of Conduct and a complete process for handling reports of violations is in development. Until it is available, please bring concerns to us by contacting a member of the HathiTrust staff or by emailing conduct-reports@hathitrust.org.
Technology overview

- Zoom Meeting Features
- Mute and Unmute
- Chat
- Speaker View and Gallery View
- Automated Transcript/Closed Captions
- Support
Mute / Unmute

When not speaking, please keep your Microphone muted

To turn on your microphone, click the “Unmute” button in the bottom left corner

Click the “Mute” button to then turn your microphone off
Chat

Use Chat for discussion or to ask a question

To open the Chat window, click the Chat button at the bottom of the Zoom application

By clicking the “Everyone” button in the Chat pop out window, you can select who you want to chat with
Overview of HOLLIS, the online catalog of the Harvard Library
CDI offers 2 separate HathiTrust collections for activation, both updated monthly:

HathiTrust Digital Library Full View U.S. Only

HathiTrust Digital Library

What's in HOLLIS

Harvard local collections
- Print collection
- Images
- Geospatial data
- Finding aids

Central Discovery Index
- Scholarly articles
- Academic repositories
- Open access
- HathiTrust
- Much more...

SEARCH RESULTS
HOLLIS home page with single search box
Search for "guide dogs"

Results are from different sources and presented in a single list:

- digital images
- articles
- print books, owned and not owned

Our user studies reveal that users prefer single list to a bento approach.
Search result from HathiTrust

Your dogs and mine: one hundred and fifteen pen drawings and eighteen etched portraits with text
Thorne, Diana, b. 1895
1935
OPEN ACCESS
ONLINE ACCESS

HathiTrust book viewer
"The tough guy"

"Snub nose Sally"
Other ways to get HathiTrust bibliographic data

https://www.hathitrust.org/data

OAI data feed

- "The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is a protocol used in libraries and archives for the automated delivery of structured bibliographic metadata. You can use this option to retrieve metadata in MARC21 or unqualified Dublin Core formats."

Tab-delimited Files ("Hathifiles")

- "These files include some bibliographic metadata as well as data elements unique to the HathiTrust collection. You can use this data to do some analysis of the HathiTrust collection. Libraries may find these files useful for collection management and deciding which records to retrieve from HathiTrust or building links to HathiTrust works."
Use of the API
The API is called when HOLLIS search results are rendered:

- If we do not already have online access, we check for access at HathiTrust, based on OCLC number.
- If found, we offer a link to HathiTrust.
Example of title that is available in HathiTrust as Limited Search only

HOLLIS search result with Limited Search in HathiTrust

Limited search tool in HathiTrust
API output for a given OCLC number

- JSON format
- Brief bibliographic data
- Digital object information
- Usage rights code
- Item URL
- Origin

JSON data:
```
"records": {
  "100489667": {
    "recordURL": "https://catalog.hathitrust.org/Record/100489667",
    "titles": [
      "Dogs I have known"
    ],
    "isbns": [],
    "issns": [],
    "oclcs": [
      "20582965"
    ],
    "lccns": [],
    "publishDates": [
      "1893"
    ]
  }
},
"items": [
  {
    "orig": "Harvard University",
    "fromRecord": "100489667",
    "htid": "hdl.handle.net\2027/hvd.hn67nn",
    "itemURL": "https://hdl.handle.net/2027/hvd.hn67nn",
    "rightsCode": "pd",
    "lastUpdate": "20200420",
    "enumcron": false,
    "usRightsString": "Full view"
  }
]
```
Example with copies from multiple institutions, limited search

Does not tell you whether you have Emergency Temporary Access

```
}

"items": [

{
  "orig": "University of California",
  "fromRecord": "000366851",
  "htid": "uc1.$b184850",
  "itemURL": "https:\/\/hdl.handle.net\/2027\/uc1.$b184850",
  "rightsCode": "ic",
  "lastUpdate": "20170310",
  "enumcron": false,
  "usRightsString": "Limited (search-only)"
}

{
  "orig": "University of Michigan",
  "fromRecord": "000366851",
  "htid": "mdp.39015027015828",
  "itemURL": "https:\/\/hdl.handle.net\/2027\/mdp.39015027015828",
  "rightsCode": "ic",
  "lastUpdate": "20200808",
  "enumcron": false,
  "usRightsString": "Limited (search-only)"
}
```
Emergency Temporary Access Service
Sample from overlap report from HathiTrust

<table>
<thead>
<tr>
<th>oclc</th>
<th>local_id</th>
<th>type</th>
<th>access</th>
<th>rights</th>
</tr>
</thead>
<tbody>
<tr>
<td>945584209</td>
<td>9900000000010203941</td>
<td>mono</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29254328</td>
<td>9900000000030203941</td>
<td>mono</td>
<td>deny</td>
<td>ic</td>
</tr>
<tr>
<td>424526137</td>
<td>9900000000080203941</td>
<td>mono</td>
<td></td>
<td></td>
</tr>
<tr>
<td>52188610</td>
<td>9900000000090203941</td>
<td>mono</td>
<td>deny</td>
<td>ic</td>
</tr>
<tr>
<td>244177315</td>
<td>9900000000110203941</td>
<td>multi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5314464</td>
<td>9900000000980203941</td>
<td>mono</td>
<td>allow</td>
<td>pd</td>
</tr>
<tr>
<td>82617130</td>
<td>9900000000990203941</td>
<td>mono</td>
<td></td>
<td></td>
</tr>
<tr>
<td>459410</td>
<td>9900000001160203941</td>
<td>mono</td>
<td>deny</td>
<td>ic</td>
</tr>
<tr>
<td>1285105</td>
<td>9900000001170203941</td>
<td>mono</td>
<td>allow</td>
<td>pdus</td>
</tr>
<tr>
<td>7748334</td>
<td>9900000081770203941</td>
<td>serial</td>
<td>deny</td>
<td>und</td>
</tr>
<tr>
<td>7748334</td>
<td>9900000081770203941</td>
<td>serial</td>
<td>deny</td>
<td>ic</td>
</tr>
<tr>
<td>19156568</td>
<td>9900000081780203941</td>
<td>mono</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18754067</td>
<td>9900000081820203941</td>
<td>multi</td>
<td>deny</td>
<td>ic</td>
</tr>
</tbody>
</table>
Upon receipt of initial overlap report

- **Reserves**
  - 750 matches, made big difference for students who suddenly weren't allowed in the library anymore

- **2.3 million "deny" titles**
  - Big load on system to add these. Took over 2 weeks.
  - Matching: used our system number from overlap report, rather than OCLC number, to eliminate matching errors (multiple matches, number not found...). Our overlap report was based on print disclosure from 2 years ago. If we had used OCLC number, even a small matching error rate of .5% would have meant 11,500 titles to clean-up
  - Had to split up input files to manage capacity
  - Multiple e-collections in Alma with a max 500k portfolios per collection
  - Priority order of collections: de-prefere Hathi so users won't be directed there if we have paid access (which generally allows simultaneous users)
Having links in Alma was critical for Online facet in HOLLIS

"Top-level" facet in HOLLIS
Serials... watch out

- We initially loaded portfolios for these titles not realizing the issues it would cause with article-level linking in Primo.
- Because the portfolios had no coverage information, when Primo evaluated where to send our users for best coverage of a journal, it chose HathiTrust.
- This meant that many links to full-text articles didn't take the user to full-text.
- We quickly de-activated the Hathi portfolios on our serial records.

Normally there is a coverage statement so that Primo knows where to direct a user for a given journal issue.
Constructing dynamic links to HathiTrust

- We defined a link pattern at the collection level, with the OCLC number as a parameter, and this pattern is inherited by every portfolio.
- This allows us to change the pattern if needed (which we did twice in the first 2 months).
- And if our OCLC number needs to be corrected, we can update our bib record without having to change anything else.
  - Generally this was only needed when our bib had 2 OCLC numbers.

```plaintext
IF(rft.oclcnum)

https://catalog.hathitrust.org/Search/Home?lookfor={rft.oclcnum}&type=oclc
&urlappend=;signon=swle:https://fed.huit.harvard.edu/idp/shibboleth

IF()

https://catalog.hathitrust.org/Search/Advanced
```
Importance of including authentication parameter in URL

What a user sees in HathiTrust when they are authenticated
Before we did that:

- Users had to login again at HathiTrust, identifying themselves as a Harvard affiliate
- Users were not accustomed to doing this because for most of our tools they are authenticated automatically through single sign-on
- Because they were not logged in at HathiTrust, they would see a "Limited search only link" and not know how to gain access to the book

What a user sees in HathiTrust when they are not authenticated
Public display of links

Online Access

Full text available at: HathiTrust Temporary Access Sept 2020 update (Choose Temporary Access, then Check Out)

- Step-by-step instructions
- While the Harvard Libraries are closed due to the COVID-19 pandemic, this title is temporarily available for Check Out via HathiTrust

Link text for HathiTrust access
General counsel: make physical copies unavailable

We only fulfill a scan request for these if there are pages missing in HathiTrust.

### Display of physical holdings

<table>
<thead>
<tr>
<th>BARCODE</th>
<th>TYPE</th>
<th>POLICY</th>
<th>DESCRIPTION</th>
<th>STATUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>32044100534189</td>
<td>Book</td>
<td>Temporarily prohibited</td>
<td></td>
<td>Available</td>
</tr>
</tbody>
</table>
Multi-volume works

- We loaded portfolios for multi-volume works, although often HathiTrust does not have all the volumes.
- If a user reports this, we adjust our records in Alma to make the missing volumes requestable for physical use.

HathiTrust record where only volumes 3 and 10 are available.
<table>
<thead>
<tr>
<th>Barcode</th>
<th>Location</th>
<th>Description</th>
<th>Temporary Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>RSLDX4</td>
<td>Offsite Storage -- In-library use only (HD)</td>
<td>v.11</td>
<td>No</td>
</tr>
<tr>
<td>RSLDX3</td>
<td>Virtual location: no physical use - use HathiTrust ETAS HD</td>
<td>v.10</td>
<td>Yes</td>
</tr>
<tr>
<td>RSLDX2</td>
<td>Offsite Storage -- In-library use only (HD)</td>
<td>v.9</td>
<td>No</td>
</tr>
<tr>
<td>RSLDX1</td>
<td>Offsite Storage -- In-library use only (HD)</td>
<td>v.8</td>
<td>No</td>
</tr>
<tr>
<td>RSM31R</td>
<td>Offsite Storage -- In-library use only (HD)</td>
<td>v.7</td>
<td>No</td>
</tr>
<tr>
<td>RSLDWZ</td>
<td>Offsite Storage -- In-library use only (HD)</td>
<td>v.6</td>
<td>No</td>
</tr>
<tr>
<td>RSLDWY</td>
<td>Offsite Storage -- In-library use only (HD)</td>
<td>v.5</td>
<td>No</td>
</tr>
<tr>
<td>RSLDWX</td>
<td>Offsite Storage -- In-library use only (HD)</td>
<td>v.4</td>
<td>No</td>
</tr>
<tr>
<td>RSLDWW</td>
<td>Virtual location: no physical use - use HathiTrust ETAS HD</td>
<td>v.3</td>
<td>Yes</td>
</tr>
<tr>
<td>RSLDWV</td>
<td>Offsite Storage -- In-library use only (HD)</td>
<td>v.2</td>
<td>No</td>
</tr>
<tr>
<td>RSLDWWU</td>
<td>Offsite Storage -- In-library use only (HD)</td>
<td>v.1</td>
<td>No</td>
</tr>
</tbody>
</table>

List of item records in Alma
Alma CloudApp
<table>
<thead>
<tr>
<th>Title</th>
<th>Author(s)</th>
<th>Institution</th>
<th>Publication Year</th>
<th>Language</th>
<th>ISBN</th>
<th>Keywords</th>
<th>MMS ID</th>
<th>Creation Date</th>
<th>Modification Date</th>
<th>Record number</th>
</tr>
</thead>
<tbody>
<tr>
<td>A dog's house</td>
<td>Micah Schraft</td>
<td></td>
<td></td>
<td>English</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>House your dog's comfort</td>
<td>Association of College Schools of Architecture</td>
<td>Washington, D.C.</td>
<td>1991</td>
<td>English</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A dog's mission</td>
<td>Harriet Beecher Stowe</td>
<td></td>
<td></td>
<td>English</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Cloud app menu in upper right, showing HathiTrust Availability as an installed app
Cloud app shows HathiTrust availability for each title in the Alma result list.
Alma import profile for loading portfolios

### Import Profile Details

- **Profile name**: HathiTrust Temporary Access (SYS)
- **Profile description**: 
- **Originating system**: Other
- **Import Protocol**: Upload File/s
- **Physical source format**: Delimited Text File
- **Delimiter**: Comma
- **Source format**: MARC21 Bibliographic
- **File name patterns**: overlap*.csv
- **Cross walk**: Yes
- **Target format**: MARC21 Bibliographic

### Input file sample

001
990011306150203941
990029734420203941
990093239120203941
990102531390203941
990151370270203941
990141207370203941
990151370270203941
990010664260203941
Inventory information

HathiTrust Temporary Access (SYS)

Inventory Operations
- Electronic
- Mixed
- Physical
- None

E-Book Mapping
- Delete/deactivate portfolios
- Use Community Zone portfolios
- Portfolio type
  - Standalone
  - Part of an electronic collection
- Electronic Collection
  - HathiTrust Temporary Access Part 2 (Choc)
- Service
  - Full Text
- Material type
  - Book
- Single portfolio
- Multiple portfolios

Activate resource upon load
Use Analytics to identify items

Goal: put items into temporary locations / policies in order to prohibit use

1. Identify portfolios in the collection(s)
2. Use this a feeder report into a new analysis for physical item records
3. Save the export of physical items with as many attributes as you may need later when you have to restore items, after ETAS is over
   a. E.g.: if items were in a temp loc before you changed them, you may need to restore them to that temp loc
4. Use the barcodes to create item sets in Alma
5. We created sets for each library separately, in order to retain accurate library facet in Primo
   a. For each library we also configured temporary locations for on campus vs. offsite storage, for appropriate routing if a previously checked out item was returned
Feeder report of portfolios

Important: there is a limit of how many results you can feed from one report into another. You may have to do this process iteratively

Attributes and filters:

- Bib MMS ID
- Bib Level - filter to not s (serial)
- Title
- Portfolio lifecycle - filter to In Repo
- Portfolio ID
- Availability - filter to Available
- Electronic Collection ID = (find # in Alma)
- Electronic Collection Public Name
Item analysis

- Barcode
- Temporary Physical Location In Use
- Temporary Library Code
- Temporary Location Code
- Temporary Item Policy
- Lifecycle (Item) - filter to Active
- Holding suppressed = filter to 1 (No)
- Internal Note 1
- Internal Note 2
- Internal Note 3
- Library Code
- Location Code
- Description
- Fulfillment Note
- Item Policy
- MMS Id - filter to "is based on results..."
- Title

MMS ID Filter Operator:
"is based on results of another analysis"
Use the Browse button to find your feeder report... note the horizontal scroll bar!

Browse option for finding your feeder report

Select the MMS ID field from the feeder report
Primo Studio Add-on

Customers who have written their own Primo add-ons contribute them to Primo Studio, and Studio allows other customers to easily add that code to their own site.
Primo Studio resources

- "Using Primo Studio" Ex Libris documentation
- Video: Ex Libris Feature Spotlight on Primo Studio
- Link to Primo Studio
- Primo developer channel on IGELU-ELUNA Slack group

Example of HathiTrust add-on from Primo Studio
Thank you!

Corinna Baksik
corinna_baksik@harvard.edu

- Please complete the survey
- Jump in on the conversation on the HathiTrust Community Slack: #2020-comm-week
- On social media: #HTCW20
- Use @cbaksik when referring to this session