On Implementing Hydra for Special Collections at Yale

Eric James
programmer/analyst
eric.james@yale.edu

June 9, 2015
Legacy

- JSS (Joel Sumner Smith Slavic Collection) (2010)
- YFAD (Yale Finding Aid Database) (2009)

Stack:
/tomcat/fedora/solr/gsearch/

Main issues:
Digitization workflow
Content models (fedora 3)
Arabic OCR (VERUS)
Slavic language indexing (solr)
EAD style sheets
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors/Editors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sbornik &quot;Py&quot;</td>
<td>N. A. Nekrasov</td>
</tr>
<tr>
<td>Sbornik &quot;Py&quot;</td>
<td>V. Zolotov</td>
</tr>
<tr>
<td>Sbornik kritických států o N. A. Nekrasovici</td>
<td>Sestavil V. Zolotov</td>
</tr>
<tr>
<td>Sborník pišťan bukovského naroda : A collection of Bukovinian folk</td>
<td>Sestavil A. Lenarchiewski. ( Iz materiál, dostavljenog G. I.</td>
</tr>
<tr>
<td>képtelenek ki a magyar utolsó ötletel Imperatoroskéru rugózégeknek</td>
<td>Kupchan konjuktúra.)</td>
</tr>
<tr>
<td>Sborník ruských narodnych literárskych pesen</td>
<td></td>
</tr>
<tr>
<td>Šborník ruských narodnych piščení / Sestavil M. Balakovým.</td>
<td></td>
</tr>
<tr>
<td>Šborník států, chitanykh v Odtičení ruskago štyrka Flesnovosti. Volume 1</td>
<td></td>
</tr>
<tr>
<td>Šborník států, chitanykh v Odtičení ruskago štyrka Flesnovosti. Volume 2</td>
<td></td>
</tr>
<tr>
<td>Šborník států, chitanykh v Odtičení ruskago štyrka Flesnovosti. Volume 3</td>
<td></td>
</tr>
<tr>
<td>Šborník států, chitanykh v Odtičení ruskago štyrka Flesnovosti. Volume 4</td>
<td></td>
</tr>
<tr>
<td>Šborník států, chitanykh v Odtičení ruskago štyrka Flesnovosti. Volume 5</td>
<td></td>
</tr>
<tr>
<td>Šborník států, chitanykh v Odtičení ruskago štyrka Flesnovosti. Volume 6</td>
<td></td>
</tr>
<tr>
<td>Šborník států, chitanykh v Odtičení ruskago štyrka Flesnovosti. Volume 7</td>
<td></td>
</tr>
<tr>
<td>Šborník všichniho po knihovno-literárturní úřadu za 1866 rok / Izdaní knižnici 1866 A. Churenín.</td>
<td></td>
</tr>
</tbody>
</table>
Welcome to the Yale Finding Aid Database

This database consists of finding aids for archival and manuscript materials at Yale University. Their purpose is to provide information about the creation, historical context, arrangement, and content of archival and manuscript collections as well as the information necessary for users to identify and request the portions relevant to their research. These finding aids vary in style, depth of detail, and amount of materials covered, but their common purpose is to describe and enable access to archival and manuscript collections.

Finding aids represent holdings from:

- Manuscripts and Archives (Sterling Memorial Library)
- Robert B. Haas Family Arts Library Special Collections
- Beinecke Rare Book and Manuscript Library
- Yale Divinity School Library
- Historical Library, Harvey Cushing/John Hay Whitney Medical Library
- Irwin A. Glimcher Music Library
- Visual Resources Collection
- Yale Center for British Art, Department of Rare Books and Manuscripts
- The Lewis Walpole Library
- Yale Peabody Museum of Natural History

For more information on conducting research with archival and manuscript materials at Yale, see Primary Sources at Yale.

Medieval and Renaissance Manuscripts at the Beinecke Rare Book and Manuscript Library are not searchable in a separate database.

Requesting Materials

Requests to view materials in the reading rooms of the Beinecke Library and Manuscripts and Archives must be submitted through this online system. Links to the request system are found at the box level in each Beinecke Library and Manuscripts and Archives finding aid. Requests to view materials at the Divinity Library and the Arts Library must be submitted through separate online forms: Divinity Library request form, Arts Library request form. To request materials at other special collections repositories please use the contact information listed in each finding aid.
Current hydra(blacklight) 1.2^6 objects

Yale University Library Digital Collections

Welcome to the new Yale University Library Digital Collections repository.
This interface will ultimately replace the many individual digital collections now available from the Yale University Library. During this period of development, we will add both new collections and new features and are very interested in your feedback.

These collections are currently in the repository.

Arabic and Persian Medical Books and Manuscripts
Arabic and Persian manuscripts and books, as well as early translations of Arabic and Persian works dating from 1300 to 1921, from the Medical Historical Library, Cushing/Whitney Medical Library.

Connecticut Sanborn Fire Insurance Maps
Over 6,500 maps in 470 atlases for Connecticut towns dating from 1880 to 1970 at 1:600 scale. These highly detailed maps provide building, outbuilding, and property footprints, labeled streets, addresses, and information about building materials and construction features for urban and residential areas of Connecticut.

Day Missions Collection: Annual Reports
Annual reports of mission agencies and institutions document educational, medical, and religious work worldwide. Dating primarily from the 1850s through the 1930s, these reports form part of the Yale Divinity Library’s Day Missions Collection, the preeminent North American collection for documentation of the history of missions and world Christianity.

Day Missions Collection: Periodicals
Periodicals published by mission agencies and worldwide religious organizations provide documentation of educational, medical, and religious work, and eyewitness accounts of events and conditions. These periodicals form part of the Yale Divinity Library’s Day Missions Collection, the preeminent North American collection for documentation of the history of missions and world Christianity.

Israel Sack Furniture Archive
The Israel Sack Furniture Archive is a comprehensive digital database of American decorative arts, consisting of over 7,000 records cataloging the material objects bought and sold by the firm of Israel Sack, Inc. As the premier antiques firm representing early American furniture for much of the twentieth century, Israel Sack, Inc., was an influential corporation, operating from 1905 to 2002 in Boston and then New York City. The Israel Sack Furniture Archive will provide students, scholars and decorative arts enthusiasts alike with an unparalleled resource of comparative material for the study of early American furniture.

Lewis Walpole Digital Images Collection
Highlights from the Lewis Walpole Library's eighteenth-century British collections: visual caricatures and satires, Horace Walpole and Strawberry Hill—views, contents, and extra-illustrated copies of Walpole’s “Description of the Villa”; selected topographical, portrait, and historical prints and drawings; ephemera such as trade cards, advertisements, bookplates, and playbills, and more.

Maurice Durand Han Nom Handwritten and Woodblock Manuscripts
Original woodblock or brush ink texts and translations of Han Nom texts into modern Romanized Vietnamese collected by Maurice Durand, a prominent Vietnamese/French scholar of Han Nom from the mid-20th century. Han Nom script uses classical Chinese characters to represent Sino-Vietnamese vocabulary and some native Vietnamese words, while other words are represented using locally created characters based on the Chinese model.

Persian Philological Texts
Selected older Persian philology texts which originate from South Asia, are rare European translations, or are reprints. All of the selected items are in the "endangered" preservation category. Many have extensive marginalia from scholars, or are from the personal libraries of Yale scholars such as Edward Salisbury.
Ingest
Ingest

- Rake task – constantly polling for content in ladybird queue table (hydra_publish and child table hydra_publish_path)
- These tables have properties, timestamps, and file locations
- The metadata files (desc, access, rights) and content (binaries) ingested from mounted disk
- Uses content model (Simple, ComplexChild, ComplexParent, Complex childUnstruct)
Ingest lessons learned

- Able to run concurrently, 1->5 instances improved throughput from 1.93 to 0.77 sec/obj
- Concurrency required use of stored procedures with SQL insert transfer rather than use of SQL updates due to locking issues
Ingest lessons learned

- Hydra_publish table property proliferation (viewOpt, ingestServer, handle, priority, attempts, hierarchyLevel, # of digital children)
- Frequent metadata updates – a pain point - mistakes, metadata schema changes (1 ex: ISO dates for date slider facet)
Ingest lessons learned

- Errors happen
- Use database error table for quick lookup
- Use well labeled and concise logging (grep is your friend)
Ingest lessons learned

- Pluggable conditional workflow sequences,
- Quick turnaround to add features such as handles, and OCR solr fieldtype conditionals
Contextual Navigation

- Scale of Henry Kissinger Papers (13000 containers, 7 layers)
- Breadcrumbs
- Context tree
- Search within
Notes on John F. Kennedy administration

The material in this folder is open for research use only with permission. Researchers who wish to gain access can fill out this form.

Published/Created:
Circa 1960-circa 1963

Series:
Series I

Box:
174

Folder:
7

Format:
Mixed Material

Rights:
Information about Access: The papers are open with permission. Ownership & Copyright: Copyright is retained by the creator of this collection for unpublished works they have authored or otherwise produced. After the lifetime of the creator, copyright passes to Yale University. Copyright status for other collection materials is unknown. Transmission or reproduction of materials protected by U.S. Copyright Law (Title 17, U.S.C.) beyond that allowed by fair use requires the written permission of the copyright owners. Works not in the public domain cannot be commercially exploited without permission of the copyright owners. Responsibility for any use rests exclusively with the user.

Source Creator:
Henry A. Kissinger papers, part II

Call Number:
MS 1981

Yale Collection:
Manuscripts and Archives

Digital Collection:
Henry A. Kissinger Papers

Citation:
Context tree

- Javascript jstree implementation
- Backed by web service within hydra that leverages solr to create json nexting
- AUTH/Z baked in for filtering selective material
- Lazy loading (chunking via toplevel, direct selection, sibling and hierarchy context supplementation, and blocks
Breadcrumbs and search within

- 2 fields directly indexed leveraging hierarchical relationships
- Breadcrumbs (component titles and links)
- Hierarchy (space separated list of PIDs going down the hierarchy ending in a wildcard)
- So “digcoll:parent digcoll:child*” is used as a filter to search within grandchildren like “digcoll:parent digcoll:child digcoll:grandchildX”
Full text search

- Default access
- Full access
- Selected access

- Default – search in `solr fulltext_open` field
- Full – search in `solr fulltext_open AND fulltext_restricted` fields
- Selected – search in `solr fulltext_open OR (fulltext_restricted AND (folder PID whitelist))`
Image Viewer

- FAIL: riiif openseadragon (slow and required caching maintenance)
- Jpegs were satisfactory in terms of resolution and zoom
- Home grown image server exposing images exposed by fedora 3 REST API
- Thumbnail in search results page
- Thumbnail strip on show page
- Single image page w/ ocr (on/off)
- Fulltext (all folder content displayed vertically)
- Thumbnails (all folder content as thumbnails)
- PDF download
- Component level AUTH/Z (thumbnail, jpg, ocr, metadata, PDF)
International Platform Association (IPA), Aug 2, 1973 [3 of 3]

Published/Created: 1973
Series: Series II
Box: 362
Folder: 4
Format: Mixed Material

Rights: Information about Access: The papers are open to research. Ownership & Copyright: Copyright is retained by the creator of this collection for unpublished works they have authored or otherwise produced. After the lifetime of the creator, copyright passes to Yale University. Copyright status for other collection materials is unknown. Transmission or reproduction of materials protected by U.S. Copyright Law (Title 17, U.S.C.) beyond that allowed by fair use requires the written permission of the copyright owners. Works not in the public domain cannot be commercially exploited without permission of the copyright owners. Responsibility for any use rests exclusively with the user.

Source Creator:
Henry A. Kissinger papers, part II

Call Number:
MS 1061

Yale Collection:
Manuscripts and Archives

Digital Collection:
Henry A. Kissinger Papers

Citation:
FOR IMMEDIATE RELEASE
AUGUST 2, 1973

AS DELIVERED PURSUANT TO THE INSTRUCTIONS OF H.R. H. R. H. OF HIS EXCELLENCY THE HONORABLE DIPLOMAT OF THE INTERNATIONAL PLATFORM ASSOCIATION.

Dr. Kissinger: We have an important dinner tonight, and we are all going to be very happy together.

Mr. Kissinger: I am very grateful for the honor that has been bestowed on us this evening, especially by the hand of Mrs. Longworth. She is a true friend of our family and one of my dearest friends here. At least that was my first date in Washington.

The second night that President Nixon was at the White House, he gave a private dinner, and we were all very happy together. Mrs. Longworth said to me, "You will do all right. You don't have any more trouble tonight." (Laughter.) She was correct.

Now, this speech represents something of an encouragement to me than to me and to everyone for not exactly overwhelmed by expressions of appreciation. More importantly, she avoids being the same old speech. Therefore, I have chosen to select a new theme and to select a wide variety of expressions of appreciation. Indeed, does not make sense in the program's history.

I have already expressed how delighted I have been to be remembered as one of the great leaders of our time, Mrs. Longworth. The Honorable Longworth, who has brought grace and dignity to this nation, is a leader of the great leaders who make history a worthwhile and exciting experience.

Last week, while I was preparing my remarks for this occasion, I opened some time looking for an appropriate Roosevelt speech. One particularly struck a receptive chord, since I have not seen many speeches before this.
FOR IMMEDIATE RELEASE

AUGUST 2, 1973

AS DELIVERED REMARKS OF DR. HENRY A. KISSINGER
ASSISTANT TO THE PRESIDENT
FOR NATIONAL SECURITY AFFAIRS
INTERNATIONAL PLATFORM ASSOCIATION
SHERATON-PARK HOTEL

8:00 P.M. EDT

DR. KISSINGER: Mr. Roosevelt, Mrs. Longworth, ladies and gentlemen:

I am very grateful for the honor that has been bestowed on me this evening, especially by the hand of Mrs. Longworth, who, it is true, is, if I may say so, one of my girl friends here. At least she was my first date in Washington.

The second night that President Nixon was at the White House, he gave a private dinner, and we were all still very stage struck, and he said, "Don't bring anyone, I have a date for you," and that date was Mrs. Longworth.

At that time, I thought of myself as a very serious professor, and during the course of the evening, Mrs. Longworth said to me, "You will do all right. You don't take yourself too seriously." (Laughter) She was wrong about the second part.
AUTH/N SSO (openidconnect OAUTH2)

Login

Local Login

Username

Password

Remember My Login

Login

Register

External Login

Log In with CAS

Google
Component level AUTH/Z datastream

```xml
<schema>
  <object sourceID="11775674">
    <rule type="Open Access" code="thumbnail"/>
    <rule type="Open Access" code="thumbnail_dl"/>
    <rule type="Open Access" code="jpg"/>
    <rule type="Open Access" code="jpg_dl"/>
    <rule type="AD Group Restriction" code="ntf">
      <item>LibraryLadybirdAdmins</item>
    </rule>
    <rule type="AD Group Restriction" code="ntf_dl">
      <item>LibraryLadybirdAdmins</item>
    </rule>
    <rule type="Open Access" code="jp2"/>
    <rule type="Open Access" code="jp2_dl"/>
    <rule type="Open Access" code="pdf"/>
    <rule type="Open Access" code="pdf_dl"/>
    <rule type="Open Access" code="oer"/>
    <rule type="Open Access" code="oer_dl"/>
    <rule type="Open Access" code="undefined"/>
    <rule type="Open Access" code="undefined_dl"/>
    <rule type="Open Access" code="metadata"/>
    <rule type="Open Access" code="metadata_dl"/>
    <rule type="Open Access" code="premis"/>
    <rule type="Open Access" code="premis_dl"/>
    <rule type="Open Access" code="cad"/>
    <rule type="Open Access" code="cad_dl"/>
    <rule type="AD Group Restriction" code="ntf_master">
      <item>LibraryLadybirdAdmins</item>
    </rule>
    <rule type="AD Group Restriction" code="ntf_master_dl">
      <item>LibraryLadybirdAdmins</item>
    </rule>
  </object>
</schema>
```
AUTH/Z flow and restriction types

- Check_user_session (verifies email, session, IP)

Check object AUTH/Z datastream (w/ PID, and component)
- OpenAccess
- Yale Only (netid or IP range)
- IP Restriction (IP on a list for object)
- NetID Restriction (netid on a list for object)
- AD Group Restriction (AD group on a list for object)
- AeonRegistration*
1. User granted permission to certain folders of digital content
2. Upon user login, an aeon AUTH/Z endpoint is called that returns JSON with PID of whitelisted folders
3. This JSON content is persisted to an aeon_assets table
4. When AUTH/Z occurs for a component of an object with type “Aeon Registration”, the aeon_assets table is checked for permissions related to user