Digital Repository Development at Yale University Library

Michael Dula
CTO, Yale University Library

December 8, 2014
Four groups comprise Library IT:

- Workstation and Technology Services
- User Experience
- Enterprise Systems and Architecture
- Digital Library & Programming Services
What do Libraries Care About?

- Metadata standards
- Discoverability
- Managing access and permissions
- Preservation
Historically, how are we doing?

- Metadata standards: inconsistent
Historically, how are we doing?

- Metadata standards: inconsistent
- Discoverability: spotty
Historically, how are we doing?

- Metadata standards: inconsistent
- Discoverability: spotty
- Managing access and permissions: several methods, none very granular
Historically, how are we doing?

- Metadata standards: inconsistent
- Discoverability: spotty
- Managing access and permissions: several methods, none very granular
- Preservation: we have duplicate copies—somewhere—that’s enough, right?
How did we get into this fix?

Walled gardens: with unicorns in them
How did we get into this fix?

LOTS of gardens, and rooms, and attics...
As of 2014...

Legacy Collection Technologies

- .Net 9
- Fedora 5
- Classic ASP 9
- PHP 4
- Greenstone 16
- ContentDM 15
- JSP 3
- A/V 3
- Other 16
By End of 2015...

Collection Technologies 2015

Fedora

Fedora 4

A/V

ContentDm

Other
How do we get from here to there?

Legacy Collection Technologies

- .Net 9
- Fedora 5
- Classic ASP 9
- PHP 4
- JSP 3
- ContentDM 15
- Greenstone 16
- Other 16

Collection Technologies 2015

- Fedora
- Fedora 4
- A/V
- ContentDM
- Other
A Note About Legacy Migrations

Metadata often turns out to be:

- **Imperfect.** *Strategy: If there is an easy bulk transformation, do it, but other fixes can be post-migration.*
A Note About Legacy Migrations

Metadata often turns out to be:

- Imperfect. Strategy: If there is an easy bulk transformation, do it, but other fixes can be post-migration.

- Not meeting minimum standards for display. Strategy: Fix or retire collection.
Metadata often turns out to be:

- Imperfect. Strategy: If there is an easy bulk transformation, do it, but other fixes can be post-migration.

- Not meeting minimum standards for display. Strategy: Fix or retire collection.

- Highly custom. Strategy: Migrate the data, but may not have a use for it immediately.
About Hydra/Fedora

- Open-source project
- Provides a platform for digital preservation and presentation
- Over 50 Fedora Members contributing financially; Yale is one of these
- Yale is also a Fedora development partner, and YUL’s Manager of Digital Library & Programming Services serves on the Fedora Leadership Committee
- Currently actively engaged in development of Fedora 4
Began in 2008 as collaboration between Stanford, UVA, Univ. of Hull, and Fedora Commons

YUL joined in 2013 as 18th member. Membership now up to around 25—recent additions include Princeton, Cornell, Case Western

“If you want to go fast, go alone. If you want to go far, go together.”
Simplified View of Yale’s Hydra/Fedora

Hydra-Heads (future)
Creating and managing objects

Blacklight
Discovering and viewing objects

Ladybird
(Yale’s Cataloging Tool)

Fedora
(Preservation)

Solr
(Index)
Initial Hydra Projects at YUL

- 9 pilot collections: http://findit.library.yale.edu
- Henry Kissinger Papers
- Migrations of about 80 legacy collections
- Research Data pilot with the Institution for Social and Policy Studies
The Findit Blacklight interface

You searched for: Digital Collection: Lewis Walpole Library

   Creator: O’Neale, Jefferyes Hamett, -1801, artist
   Sumpter, Edward, fl. 1783-1787, publisher
   Published/Created: [London] : [Edward Sumpter], [ca. 1763]
   Digital Collection: Lewis Walpole Library

2. *Drawings faithfully copied from nature at Naples : and with permission dedicated to the Right Honourable Sir William Hamilton, His Britannic Majesty’s Envoy Extraordinary and Plenipotentiary at the court of Naples / by his most humble servant Frederick Rehberg, historical painter in His Prussian Majesty’s service at Rome, MDCCXCIV.*
   Creator: Pirro, Tommaso, ca. 1752-1824, associated name
   Rehberg, Friedrich, 1758-1835
   Published/Created: [London] : Publish’d October 12th, 1797, by S.W. Fores, No. 50 Piccadilly, corner of Sackville Street. Prints & drawings lent out on the plan of a circulating library, [October 12th, 1797]
   Digital Collection: Lewis Walpole Library
Quicksearch Interface (http://search.library.yale.edu)

Yale University Library

BETA
Quicksearch

Catalogs

The cooperative apartment house: a reference guide
- Author: Brett, Wyckoff, Potter, Hamilton
- Published: New York, [1937]
- Location: Law Library SS 18a no.13
- Format: Books

New York, New York: how the apartment house transformed the life of the city (1869-1930)
- Author: Hawes, Elizabeth, 1940-
- Published: New York : A.A. Knopf, 1993
- Location: Law Library NAY860 .H34 1993
- Format: Books

Apartment 3A
- Author: Daniels, Jeff
- Location: HAAS ARTS LIBRARY, Drama Collection 61 Day (1rd) PS3554 .A65 2000 (LC)
- Format: Books

The apartment
- Published: Culver City, CA : Metro-Goldwyn-Mayer ; Beverly Hills, CA : Distributed by Twentieth Century Fox Home Entertainment, ©2007
- Location: Bass Library, Yale College Council DVD Collection
- Format: Video

Articles

Gonzalez, Evelyn Diaz
- Citation: The Bronx, 2004, p. 83
- Format: Available Online

New York Times (1857-1922), 9/6/1914, ISSN: 0362-4331, p. RD1
- Format: Full Text Available

New York Times (1857-1922), 7/19/1914, ISSN: 0362-4331, p. RD1
- Format: Full Text Available

New York Times (1857-1922), 7/26/1914, ISSN: 0362-4331, p. RD1
- Format: Full Text Available

- Format: Full Text Available

PLANS PROVIDE AIRY APARTMENTS
Immediate Priorities: Scale and Security

- We have to get large—1 PB already queued up for ingest and/or migration

- We have to provide multiple levels of security, from Open Access to highly mediated
What’s next?
Possible Future Directions

▪ Curated Research Data
▪ Self-Archiving by Yale community
▪ Streaming A/V Support
▪ Online exhibitions
▪ Active Preservation tools
▪ Advanced Interfaces: GIS, Digital Humanities, Data Visualization
Research Data

- Pilot project underway with ISPS
  - Learn how to ingest research studies as complex objects
  - Preserve them and provide permanent links
  - How do they display in our standard Blacklight interface?

- Develop custom interfaces around this content type

- Integration with other systems
Evidence-Based Practice: True or False? [poster for ACRL 2013 Conference]

**Descriptions**

<table>
<thead>
<tr>
<th>Resource type</th>
<th>Poster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Evidence-Based Practice: True or False? [poster for ACRL 2013 Conference]</td>
</tr>
<tr>
<td>Creator</td>
<td>Nancy E. Adams</td>
</tr>
<tr>
<td>Contributor</td>
<td></td>
</tr>
<tr>
<td>Description</td>
<td>Briefly describes foundational concepts in evidence-based practice (EBP) and addresses epistemological assumptions of EBP.</td>
</tr>
<tr>
<td>Keyword</td>
<td>evidence-based practice, information literacy, Association for College and Research Libraries (ACRL), academic libraries</td>
</tr>
<tr>
<td>Rights</td>
<td>Attribution-NonCommercial-NoDerivs 3.0 United States</td>
</tr>
<tr>
<td>Publisher</td>
<td></td>
</tr>
<tr>
<td>Date Created</td>
<td>March 2013</td>
</tr>
<tr>
<td>Subject</td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td>English</td>
</tr>
</tbody>
</table>
What About A/V Materials?

http://www.avalonmediasystem.org/project
Support for Online Exhibitions

Digital Preservation Services

- Bit Preservation
- Secure Storage with Managed Access
- Obsolescence Monitoring
- Provenance and Authenticity Assurance
- Standards Compliance
- Format migration and emulation services
Possibilities for Advanced Interfaces

- GIS
- Data manipulation and visualization
- Digital Humanities
How can we do all this at speed and at scale?

It took 300 years for the Yale Library catalogs to grow to around 10 million items. Just ONE large digital collection that we are currently working on will contain about 10 million items.
Questions?

Michael Dula