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
- Classic ASP: 9
- JSP: 3
- A/V: 3
- Fedora: 5
- PHP: 4
- ContentDM: 15
- Greenstone: 16
- Other: 16
How do we get from here to there?
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.**
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.

- 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.”
Yale's Technology Stack

Hydra-Head
- Creating and managing objects (CRUD)
  - Hydra Model Logic

Blacklight
- Discovering and viewing objects (R)
  - Search/Facet Logic

Hydra Access Controls

Ladybird
- Yale's Cataloging Tool

Solrizer

Hydra Models

Fedora
- (Preservation)

Solr
- (Index)

Managed Storage

Bookreader
- Complex Object Display

Downloadable PDF

Single Image Zoom

Bookreader
- Complex Object Display

Hydra Model Logic

Image Request

Image Retrieval

Media Server

Link to Images

Yale University Library
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
Quicksearch Interface (http://search.library.yale.edu)
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
- 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

Resource type: Poster
Title: Evidence-Based Practice: True or False? [poster for ACRL 2013 Conference]
Creator: Nancy E. Adams

Description: Briefly describes foundational concepts in evidence-based practice (EBP) and addresses epistemological assumptions of EBP.

Keyword: evidence-based practice
information literacy
Association for College and Research Libraries (ACRL)
academic libraries

Rights: Attribution-NonCommercial-NoDerivs 3.0 United States

Publisher: [Publisher Name]
Date Created: March 2013
Subject: [Subject]
Language: English

Yale University Library
What About A/V Materials?

http://www.avalonmediasystem.org/project

Media Collections Online

Browse by

- Format
  - Moving Image (360)
  - Sound Recording (9)
- Date
- Genres
- Collection
- Unit

Featured Film

- The Masters of Disaster (1985)

Featured Film Collection

- Indiana University Libraries Moving Image Archive

Media Collections Online houses audio and video collections.

- Use the search box, or browse using the terms on the left side to discover content.
- Some content is only available when you sign in with your ID and passphrase.
- Login to Oncourse or Canvas to access course materials.
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