Hypatia
Hydra Platform for Access to Information in Archives

DLF Forum * Baltimore * October 31, 2011

Stanford University
Bradley Daigle
Julie Meloni
Naomi Dushay

University of Virginia
Tom Cramer
Michael Olson
Naomi Dushay

Stanford University
Introduction

AIMS

Born Digital Materials (& Forensics)

Hypatia Functional Requirements

Data Models & Loading

Demonstration

Q&A

Discussion

In Sum & Looking Forward

Tom Cramer

Bradley Daigle

Michael Olson

Naomi Dushay

Julie Meloni & Michael Olson

Discussion

Tom Cramer
What is Hypatia?

• **Hydra Platform for Access to Information in Archives**
• Repository-powered solution for digital archival materials management, preservation and access
• One component in a larger (eco)system for archivists
• Open source software based on Hydra & Fedora
• The potential nucleus of a larger, sustained, collaborative effort
Origins

• Outgrowth of AIMS project
• Leveraging the Hydra project
• Functional requirements and content from the AIMS partners (Virginia, Hull, Stanford, Yale)
• Technical Development by Stanford, Virginia & MediaShelf (contract)
What is Hydra?

**Partners**
- DuraSpace
- Northwestern University
- Notre Dame
- Rock & Roll Hall of Fame
- Stanford University
- University of Hull
- University of Virginia

+ half dozen more in ramp-up mode

**Technology**
OSS stack featuring Fedora, solr, Ruby on Rails, Blacklight

“Solution Bundles”
- IR
- ETD’s
- Research Data
- Video
- Images

• Archives → Hypatia
• Open Access Articles

- Digitization Workflow
- Digital Monograph Acquisitions
- Exhibits
- Digital Preservation

Hypatia Partners
- Stanford University
- Yale University
- University of Hull
- University of Virginia
The Workflow

Physical Materials → EAD → Repository

Born Digital Materials → Forensic Extraction & Processing → Discovery & Access

Object Management & Preservation → Arrangement & Description

Hypatia Partners
Stanford University
University of Hull
Yale University
University of Virginia
Iterations & Enrichment

EAD Enrichment: IDs and URLs for files / containers

Two Phase Data Processing: Reprocess for object-level access

Hypatia Partners
Stanford University
University of Hull
Yale University
University of Virginia
AIMS: BORN-DIGITAL COLLECTIONS

An Inter-Institutional Model for Stewardship (AIMS)

October 2009-September 2011

Funded by The Andrew W. Mellon Foundation
The Purpose

- To bring four top research libraries together to develop a common framework for the collecting, organizing, preserving and delivering of born-digital archival collections.
AIMS: the UnConference

Not two full weeks into my new job as Digital Archivist at UVA on the AIMS grant, I rolled up my sleeves to facilitate and host an unconference with my fellow Digital Archivists. Our unconference would be two full days of discussions, demonstrations, lightning talks, and networking with digital archivists from around the globe. At first the thought was a little terrifying - I'm not even fully sure I know what this job is yet, how could I actually lead discussions on the salient topics? But my fears were baseless; all the unconference attendees were thoughtful, articulate, and lively participants. I learned much more from them than they probably did from me.

The unconference was held on the 13th and 14th of May at the Omni Hotel in Charlottesville. The 27 participants represented libraries, archives, museums, and digital humanities centers across the US, Canada, and the UK. Despite the differences in our institutions, backgrounds, and training, we learned that we not only shared similar challenges, but also the same hopes for collaboration and innovation.

The first day started off with a round of lightning talks. Each participant had 5 minutes to present a topic, project, problem or idea that they were interested in talking about. The variety in the talks was remarkable to me; traversing the breadth and depth of all that can be thought of as “born-digital” and the many processes involved in managing it. The lightning talks were also great way to get an introduction to each participant, as well as their perspective or the particular issues they were dealing with in their institutions. A brief outline of each of the talks is available on the AIMS Unconference Wiki.
Outcomes

- Energized collection development & delivery of born-digital archives
- Portable tool set for the management of born-digital collections that other institutions could adapt
- A foundation on which to build a digital archivist community
- Published guide of methods used and lessons learned
Key to Success

- AIMS is based on stewardship of materials—both analogue and digital
- Informs development of tools and solutions that are informed by archival practice
- Identify the various stages in workflows and suggest methodological solutions
Follow us

- Blog: Born Digital Archives
  http://born-digital-archives.blogspot.com

- Website: AIMS – Born Digital Collections
  http://www2.lib.virginia.edu/aims
Functional Requirements Gathering

• Created by AIMS Digital Archivists’
• January – March 2011
• Initial Focus on Arrangement and Description – a tool for Archivists’
• Second focus on Discovery and Access
Functional Requirements for Arrangement and Description

ARD 00: Fundamentals (MM)
ARD 01: Graphical User Interface (MM)
  Overview
  Questions
    ARD 01.01: Representation and manipulation of hierarchy
    ARD 01.02: Drag and drop functionality
    ARD 01.03: Sort records
    ARD 01.04: Copy and paste of hierarchical structure
ARD 02: Technical Metadata (PC)
  Overview
    ARD 02.01: File-level technical metadata
    ARD 02.02: Directory-level technical metadata
    ARD 02.03: Presentation of technical metadata
ARD 03: Descriptive Metadata (SW)
  Overview
    ARD 03.01: Importing existing EAD
    ARD 03.02: Viewing/editing descriptive metadata
    ARD 03.03: Creating new description and intellectual arrangement
    ARD 03.04: Exporting EAD data
    ARD 03.05: Controlled vocabularies
ARD 04: Rights/Restrictions (MM)
  Overview
    ARD 04.01: Date-based access restrictions with automatic removal
    ARD 04.02: Access restrictions to be removed manually at a later date
Functional Requirements for Discovery and Access in AIMS Hypatia
(Hydra Platform for Access To Information in Archives)

Will need a publication method that moves files to D&A environment

- D&A_01: Filter Content
- D&A_02: Descriptive Metadata (import)
- D&A_03: Administrative view
- D&A_04: Search
- D&A_05: Display / View Files / Representation (text, image, audio, video, web, etc.)
- D&A_06: Registration & Terms & Conditions
- D&A_07: Sort on field (NEW)
- D&A_07B: Select (individual or multiple, check or highlight)
- D&A_08: Reporting (MM)
- D&A_09: Edit / Annotate
- D&A_10: Request (mediated v easy download)
- D&A_11: Upload / Edit
- D&A_12: Browse

D&A_01: Filter Content (ge)
How do we get to Hypatia 1

- Archival Description – Encoded Archival Description (EAD)
  - Repository data
  - Collection data (title, extent, …..)
  - Physical and Intellectual arrangement

Challenges with EAD
- Encoding Standards institutional specific
- Doesn’t scale for born digital archives (100,000 of files)
How do we to Hypatia 2

- Archival payload – disk images / files
  - Typically stored on obsolete media
  - Minimal descriptive metadata
The POWER of Digital Forensics

• Specialized software to help archivists’ preserve provenance by:
  • Migrating data off of legacy at risk media
  • Captures create, modify date, last accessed date
  • Preserves original media file paths, OS and low level formatting?
  • Original applications including fonts that created the data
Digital Forensic Processing

- Archivists use commercial or open source software to tag large quantities of born digital archival materials
  - Keyword, pattern search to find files that have sensitive information (Health records, Credit Card data, etc.)
- Bulk edit tagging for restricted files, subject, source media (what disk did the file come from?)
hydra = blacklight
<table>
<thead>
<tr>
<th>Metadata Type</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive Metadata (MODS)</td>
<td></td>
</tr>
<tr>
<td>Content Metadata (Stanford)</td>
<td></td>
</tr>
<tr>
<td>Rights Metadata (hydra)</td>
<td></td>
</tr>
<tr>
<td>RELS-EXT (model, parent, ...) (Fedora)</td>
<td></td>
</tr>
<tr>
<td>DC (Fedora)</td>
<td></td>
</tr>
</tbody>
</table>
EAD

Digital Content

Digital Content

Digital Content

Digital Content

Digital Content

Digital Content

Digital Content

Digital Content

Digital Content
FTK Output not designed for this:

FTK practices vary
Whole (digital) archival management

• A Case Study: Feigenbaum Papers at Stanford, prior to SALT
• Previously:
  - Files on a separate file store
    - Permissions management & preservation challenges
  - A distinct index with its own faceted browser (Flamenco)
  - A separate Drupal site for collection landing page
  - A separate Mysql db for tags
  - A separate Finding Aid, stored in Archivists Toolkit
• = a Nightmare to synchronize, update and migrate
Hypatia Benefits

• Integrated solution for archival digital objects management & access
• Granular permissions management for discover, read, edit, and administrate
• Support for multiple arrangements: physical, logical, archival
• Enables ongoing processing as resources become available
• Integrated approach for digital preservation
WIIFM?

• Open source code base for digital archives management
• Nucleus for further, community development
• Forensic toolkit patterns, best practices, Fedora loading scripts
• Data models for EAD and digital archival objects in Fedora
• Functional requirements for arrangement, description, discovery and access for digital archives
Next Steps

• Pilot usage in a archives processing digital materials
• Another round of development
• Development of bulk permissions management, arrangement & description
• Experiment with UI and tools for archivists and for end users
80/20 – 8 Weeks of Development

https://github.com/projecthydra/hypatia/graphs/impact
Connect

Demo: http://hypatia-demo.stanford.edu

Wiki: https://wiki.duraspace.org/display/HYPAT/Home

List: http://hypatia-tech.googlegroups.com

Code: https://github.com/projecthydra/hypatia

Hydra: http://projecthydra.org

AIMS: http://www2.lib.virginia.edu/aims/