Developing the Hydra - Blacklight Way

CNI Membership Meeting
St. Louis * April 1, 2014

Tom Cramer
Chief Technology Strategist
Stanford University Libraries
@tcramer
Agenda

• What is Hydra?
• What is Blacklight?
• Where are they going?
• What does it all mean?
What Is Hydra?

- A robust repository fronted by feature-rich, tailored applications and workflows ("heads")
  ➤ *One body, many heads*

- Collaboratively built “solution bundles” that can be adapted and modified to suit local needs.

- A community of developers and adopters extending and enhancing the core
  ➤ *If you want to go fast, go alone. If you want to go far, go together.*
Fundamental Assumption #1

No single system can provide the full range of repository-based solutions for a given institution’s needs,

...yet sustainable solutions require a common repository infrastructure.
A Tale of Two Worlds

Stephen Jay Gould

Museum of Comparative Zoology

Harvard University

Cambridge, MA 02138

Christopher Wren, the leading architect of London’s reconstruction after the great fire of 1666, lies buried beneath the floor of his most famous building, St. Paul’s Cathedral. No elaborate sarcophagus surrounds the site. Instead, we find only the famous epitaph written by his son and now inscribed into the floor: “A monument reared, a structure— if you are searching for his monument, look around. A tall gravestone, perhaps, but I have never read a finer testimony to the central importance—one might even say sanctity—of actual places, rather than replicas, symbols, or other forms of vicarious resemblance.

An odd coincidence of professional life recently turned my thoughts to this most celebrated epitaph when, for the second time, I received an office in a spot laden with history, a place still resplendent with ghosts of past events both central to our common culture and especially meaningful for my own life and choices.

In 1973, I spent an academic term as a visiting researcher at Oxford University. I received a century of office space on the upper floor of the University Museum. As I set up my books, fossil skulls and microscope, I noticed a metal plaque affixed to the wall, informing me that this reconfigured space of shelves and curricula had been, originally, the site of the most famous public confrontation in the early history of Darwinism. On this very spot, in 1860, just a few months after Darwin published The Origin of Species, T.H. Huxley had drawn his rhetorical sword, and soundly skewered the slick but superficial champions of creationism, Bishop "Snapy Sam" Wilberforce.

As with most legends, the official version ranks as mere chaff behind a much more complicated and multifaceted truth. Wilberforce and Huxley did put on a splendid, and largely spontaneous, show—but no clear victor emerged from the scuffle and Joseph Hooker. Darwin’s other champions, made a much more effective reply to the Bishop, however forgotten by history. See my essay on this debate, entitled "Knight takes Bishop!" and published in earlier volume of this series: Daily for Brontosaurus.

I can’t claim that the lingering presence of these Victorian giants intensified my resolve or improved my work, but I loved the sense of continuity evoked to me by this happy circumstance. I even traced the etymology—for "circumstance" means "standing around" in

KB) directions (in FSL format)
dwi.bvals (770 Bytes) The diffusion gradient strength (in FSL format)
Repository-Powered Approach

ETDs (Theses), Books, Articles, Images, Audio-Visual, Research Data, Maps & GIS, Documents

Scalable, Robust, Shared Management and Preservation Services
One Body, Many Heads...

ETDs (Theses)  Books, Articles  Images  Audio-Visual  Research  Maps & GIS  Documents

Scalable, Robust, Shared Management and Preservation Services
Hydra Technical Framework
CRUD in Repositories

Create/Submit/Edit (CUD)

Search/View (R)

Repository/ Persistent Storage
CRUD in Repositories

Create/Submit/Edit (CUD)

Search/View (R)

Repository/ Persistent Storage
Major Hydra Components

- hydra-head
- Fedora
- Solr
- Solrizer
- Blacklight (Read Only)
- Rails Plugin (CUD)
A Note on Ruby on Rails

- *Rapid* application development for web applications: “Convention over configuration”
  - 10x productivity
- **Supportable:** MVC (Model-View-Controller) and Rails framework make code well-structured, predictable
- **Testable:** Rspec and Cucumber give powerful, automatable, testing tools
- **Learnable:** Stanford went from 1 to 8 Ruby savvy developers in one year (no new hires)
  - 1 week learning curve to basic proficiency
A Note on Fedora

• Flexible, Extensible, *Durable* Object Repository Architecture
  – Flexible: model and store any content types
  – Extensible: easy to augment with apps and services
  – Durable: foundation of preservation repository

• Proven, sustained and successful digital repository
  – 100’s of adopters; 12 years of development, 3 releases
  – Vibrant community & funding under DuraSpace

• Fedora 4 in development; co-evolving with Hydra
Hydra Community Framework
Fundamental Assumption #2

No single institution can resource the development of a full range of solutions on its own,

...yet each needs the flexibility to tailor solutions to local demands and workflows.
Hydra Philosophy – Community

• An open architecture, with many contributors to a common core

• Collaboratively built “solution bundles” that can be adapted and modified to suit local needs

• A community of developers and adopters extending and enhancing the core

• “If you want to go fast, go alone. If you want to go far, go together.”

   One body, many heads
Community

• Conceived & executed as a distributed, collaborative, open source effort from the start

• Initially a joint development project between Stanford, Univ of Virginia, and Univ of Hull

• Hydra Partners are the backbone of the project
  • Coalition of the willing
  • No fees or dues
  • Apache-style consensus and governance

• Steering Group provides administration, continuity, and serves as backstop when needed
  • But no central planning, no Project Director, no “one” architect
Hydra Partners...

...are individuals, institutions, corporations or other groups that have committed to contributing to the Hydra community; they not only use the Hydra technical framework, but also add to it in at least one of many ways: code, analysis, design, support, funding, or other resources.

*Hydra Partners collectively advance the project and the community for the benefit of all participants.*

https://wiki.duraspace.org/display/hydra/Hydra+Community+Framework
Code Licensing

• All Hydra code is available under Apache License, Version 2.0

• All code commitments are being managed through Contributor License Agreements
  • Individual – so each developer is clear about what they are contributing
  • Corporate – so each institution is clear about what it is contributing

• Code contributors maintain ownership of their IP
  • ...and grant a non-exclusive license to the project and its users
Hydra Current State
Hydra Partners

OR = Open Repositories Conference
Hydra Partners and Known Users

OR = Open Repositories Conference
University of Virginia, 2008

LibDevConX^3, Stanford, 2012: (Hydranauts and Fellow Travelers)

Hydra Partners @ UCSD, 2012
First Worldwide Hydra Connect @ UCSD, 2013
In a Nutshell, Project Hydra...

...has had 7,716 commits made by 62 contributors representing 60,904 lines of code.

...has a codebase with a long source history maintained by a very large development team with stable Y-O-Y commits.

...took an estimated 15 years of effort (COCOMO model) starting with its first commit in October, 2009 ending with its most recent commit 6 days ago.
Trending

Commits per Month

Contributors per Month
Hydra Heads: Institutional Repositories

University of Hull
Hydra Heads: Institutional Repositories

ScholarSphere / Sufia @ Penn State
Hydra Heads: Institutional Repositories

Hydramata

• 6-institution collaboration
• Build a shared Hydra head to serve as an IR
• Distributed, agile team

University of Notre Dame
Northwestern University
Indiana University
University of Cincinnati
University of Virginia
Hydra Heads of Note

Avalon & HydraDAM for Media

UCSD DAMS

BPL Digital Commonwealth

Sufia

Northwestern Digital Image Lib.
Hydra-based Applications at Stanford

ETD’s – Electronic Theses & Dissertations

SALT – Self-Archiving Legacy Toolkit

EEMs – Everyday Electronic Materials

Argo – Repository Reporting and Management

Hypatia – Archives & Special Collections

SDR – Web UI (aka Hydrus)
Sustainability

No animals were harmed in the making of this film.
Current State of Hydra

- Doubled in size, thrice, since its founding
- Achieved all its original goals
- Entering a new phase
Three critical priorities

1. Technical framework continue to advance to enable easy code sharing and reuse

2. Community framework evolves to accommodate growth in partners.

3. Create a full suite of solution bundles
Eight strategic priorities

1. Develop solution bundles
2. Develop turnkey applications
3. Grow the Hydra vendor ecosystem
4. Codify a scalable training framework to fuel community growth
5. Develop a documentation framework
6. Ensure the technical framework allows code sharing
7. Refresh and intensify the community ties
8. Grow the community of users and adopters
Blacklight
Blacklight for Repositories

• Repository-agnostic, feature-rich, content-aware, turnkey access interface

• Aggregate content from multiple repositories, with links back to source systems

• Vibrant, multi-institutional, open source community on its own

• Can be used independently, or as the first component of, Hydra
Blacklight

• **Who’s Using It?**
  – Scores (probably hundreds) of institutions
  – Notables: UVa, Stanford, Columbia, Cornell, Princeton, Yale, Johns Hopkins, WGBH, BPL, NYPL, Danish Technical University, Tufts, Indiana University, Univ. of Wisconsin, et al.
  – Every Hydra institution

• **Remarkably strong community**
  – 328 members on [blacklight-development@googlegroups.com](mailto:blacklight-development@googlegroups.com)
Blacklight Development

In a Nutshell, Blacklight...

... has had 2,135 commits made by 49 contributors representing 10,855 lines of code

... has a codebase with a long source history maintained by a large development team with increasing Y-O-Y commits

... took an estimated 3 years of effort (COCOMO model) starting with its first commit in October, 2009 ending with its most recent commit 1 day ago

30 Day Summary
Mar 2 2014 — Apr 1 2014
79 Commits
6 Contributors
including 2 new contributors

12 Month Summary
Apr 1 2013 — Apr 1 2014
508 Commits
Up +152 (42%) from previous 12 months

19 Contributors
Down -2 (9%) from previous 12 months

http://www.ohloh.net/p/blacklight
CLIO at Columbia University – Bento Box results w/ article search
Forward at University of Wisconsin
The John Dugdale Collection of the Revs Institute

Total items: 693

Dugdale began his career in 1933 as a motor sports photo-journalist with Autocar magazine. In the 1960s and 1970s he held several important positions in the automotive industry, most notably as the public relations director for the Jaguar, Rover, and Triumph Division at British Leyland of America. The John Dugdale Collection consists of monochrome and color photographs from 1935 through 1948. The collection includes rare images of attempts at the Land Speed Record from Nazi-era Germany along the Autobahns as well as John Cobb’s 1939 Napier-Railton land speed car at the Bonneville Salt Flats in Utah. Other images include the French Grand Prix and German Grand Prix in 1938 and p... more »

Revs Digital Library at Stanford: Blacklight for Image Collection
French Revolution Digital Archive

A collaboration of the Stanford University Libraries and the Bibliothèque nationale de France

About the collections

The French Revolution Digital Archive (FRDA) is a multi-year collaboration of the Stanford University Libraries and the Bibliothèque nationale de France (BnF) to produce a digital version of the key research sources of the French Revolution and make them available to the international scholarly community. The archive is based around two main resources, the Archives parlementaires and a vast corpus of images first brought together in 1989 and known as the Images de la Révolution française.

Archives parlementaires / Parliamentary archives

The Archives parlementaires is a chronologically-ordered edited collection of sources on the French Revolution. It was conceived in the mid 19th century as a project to produce a definitive record of parliamentary deliberations and also includes letters, reports, speeches, and other first-hand accounts from a great variety of published and archival sources. Because of copyright limitations, FRDA contains the AP volumes covering the years 1787-1794. The text of these volumes has been marked up using TEI so that speakers, places, dates, and terms in the published index can be easily found. Users can see both scanned images of the AP pages or just the texts.

Images de la Révolution française / Images of the French Revolution

The Images are composed of high-resolution digital images of approximately 14,000 individual visual items, primarily prints, but also illustrations, medals, coins, and other objects, which display aspects of the Revolution. These materials were selected, mainly from the collections of the Département des Estampes et de la photographie, but also from other BnF departments, and include thousands of images for the important collections entitled Henri II and De Vincx. Detailed metadata exists for the images, so that researchers can search by artist, subject, genre, and place.
Digital Manuscript Index at Stanford: **Blacklight for MS Catalog**
Bassi-Veratti at Stanford: Blacklight for EAD and digital archive
Rock & Roll Hall of Fame: Blacklight for Catalog, EAD and Media
OpenVault: *Blacklight for Video at WGBH*
Digital Commonwealth at BPL: **Blacklight for statewide repository**
GeoBlacklight Infrastructure
Darren Hardy, March 2014

Search Cluster
- Solr
- WxS

Discovery Cluster
- Geo Blacklight
- Download Proxy

Delivery Cluster
- GeoServer
- WxS
- GeoWeb Cache

Storage Cluster
- PostGIS
- NFS
- Fedora Repository
- ISO/FGDC to MODS
- CRUD

Searchers
Analysts
Curators Managers

GeoBlacklight: Blacklight for geospatial resources
Maps of Africa: A Virtual Collection
A digital collection of African Maps at the Stanford University Libraries

Limit your search
Geographic Region
Topic
Era

Stanford’s Maps of Africa Collection became a major research resource in August of 2001 with the acquisition of the Oscar I. Norwich Collection of Maps of Africa and its islands, 1486 – ca. 1900. This acquisition added 318 maps to Stanford’s existing holdings of Maps of Africa. The collection was once again augmented in 2010 with the purchase of 294 maps from Caroline Batchelor, and then again in 2011 with a small set of maps from Rodney Shirley. Altogether, the Maps of Africa Collection holds over 700 maps, 632 of which are searchable and viewable through Stanford’s Image Gallery.

A Map of Africa
1793

The Lower Niger.

Cape Colony, Natal &c.

Africa from the latest Authorities
1806


A New and Accurate Map of Africa.

The Hydra-Blacklight Way
Why Hydra?

• Need a flexible, rich, robust repository solution

• Be part of a community
  – Shared resources, shared risk, shared advances

• “We” is smarter than “me”
  – Especially in this crowd

• Control your own destiny

• Build institutional capacity

• Practice best practices

• The way work is getting done now
Breadth & Depth: Access, Mgmt & Pres

ETDs (Theses)  Books, Articles  Images  Audio-Visual  Research Data  Maps & GIS  Documents

Scalable, Robust, Shared Management and Preservation Services
If You Want To Go Fast  
...go alone.
...use Hydra

1) Notre Dame deployed a video cataloging head in 6 weeks, from scratch

[Diagram image]

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

2) Hypatia development at Stanford

- First 20% of project took 80% of time (modelling, data acquisition and preparation)
- 80% of deliverables (coding) happened in 8 week spurt, by leveraging Hydra
Becoming a Hub

• Working Groups forming
  – Repository Managers
  – Archivists
  – Hydra RDF
  – Fedora...

• Jobs
Sustainability...

• For institutions
  *Partners help shoulder the load*

• For other projects
  *Post-project maintenance & governance*

• For leveraging funders’ investments
  *Adoption by growing user community*
The Hydra Way

• Not a directed project

• Investment in a framework, not an application
  – Contributions back to core code base

• Investment in a community, not a vendor
  – Contributions back to community: Training, documentation, modeling, evangelism, support

• Travel / face-time