



A Technical and Community Framework For Customized, Shared Repository Applications

Tom Cramer
Lynn McRae
Willy Mene
Bess Sadler
Stanford University

Chris Awre
Richard Green
University of Hull

Tim Sigmon
University of Virginia

Thornton Staples
DuraSpace

Open Repositories 2010

July 6, 2010

We have a ~~problem~~ need opportunity

At Open Repositories 2009, the notion of “durable objects, ephemeral applications” was a major theme.

Repositories are well established to provide valuable infrastructure to manage and preserve digital objects.

But we need a framework to quickly produce rich, user-facing applications that interface between repositories and users.

Open Repositories 2010



July 6, 2010

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.

For instance...

An ETD solution...

- Single PDF
- With auxiliary data files
- Simple, prescribed workflow
- Integrated with student administration system
- Streamlined UI for depositors, reviewers & readers

A digitization workflow system...

- Potentially hundreds of files type per object
- Complex, branching workflow
- Sophisticated operator (back office) interfaces

A general purpose institutional repository

- Heterogeneous file types
- Simple to complex objects
- General purpose user interfaces

Distinct Application Needs

More than one dozen distinct repository application needs across three institutions.

- **Electronic theses & dissertations**
- **Open access articles**
- **Data curation application(s)**
- **General purpose institutional repository**
- **Manuscript & archival collection delivery**
- **Library materials accessioning tools**
- **Digitization workflow system**
- **And more...**

Shared, Primitive Functions

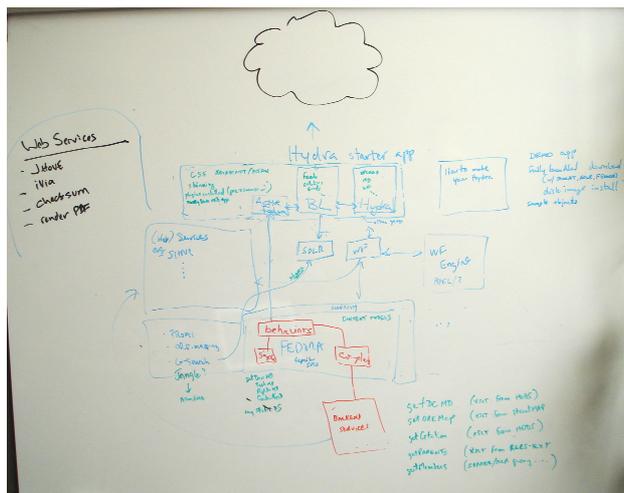
- **Deposit – uploading simple or multipart objects, singly or in bulk**
- **Manage – editing an object's content, metadata and permissions**
- **Search – full text and fielded search supporting both user discovery and administration**
- **Browse – sequential viewing of objects by collection, attribute or ad hoc filtering**
- **Deliver – viewing, downloading & disseminating objects through user and machine interfaces**

Hydra Philosophy – Technical

- Tailored applications and workflows for different content types, contexts and user interactions
- A common repository infrastructure
- Flexible, atomic data models
- Modular, “Lego brick” services
- Library of user interaction widgets
- Easily skinned UI

One body, many heads

Technical Framework



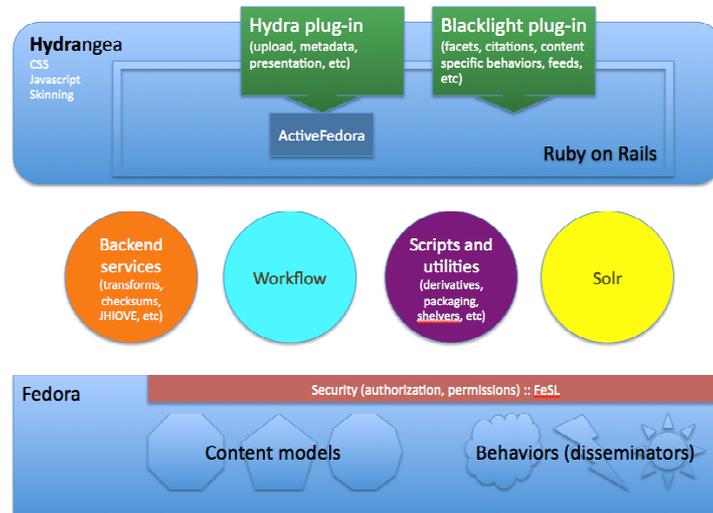
Technical Framework - Components

- **Fedora** provides a durable repository layer to support object management and persistence
- **ActiveFedora**, a Ruby gem for creating and managing Fedora objects
- **Solr**, provides fast access to indexed information
- **Blacklight**, a Ruby on Rails plugin that sits atop solr and provides faceted search & tailored views on objects

Technical Framework - Components

- **Hydra Plugin**, a Ruby on Rails library that provides create, update and delete actions against Fedora objects
- **Services**, providing discrete, reusable Lego bricks of functions (e.g., indexing, checksumming, MD transforms, etc.)
- **Hydrangea**, a web application that bundles all the RoR components, user interaction and hooks to services into a single, adaptable package

Technical Framework



Open Repositories 2010



July 6, 2010

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.

Open Repositories 2010



July 6, 2010

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 collaborative, open source effort from the start
- Joint development project among Stanford, University of Virginia, University of Hull, DuraSpace, and MediaShelf
- 20+ active contributors across 5 organizations
- Half dozen (plus?) potential adopters (& contributors!) waiting in the wings

Ultimate Objective

Hydra's ultimate objective is to effectively intertwine its technical and community threads of development, producing a community-sourced, sustainable application framework.

This framework will ideally provide rich and robust repository-powered solutions as an integrated part of an overall digital content management architecture.

Such solutions can meet the distinct needs of digital library, institutional repository, discipline repository, research, preservation and publishing workflows.

Hydra Heads

- Electronic Theses & Dissertations
- SALT (digital archival materials)
- EEMs (accessioning born digital library materials)
- IR @ Hull
- Images & Manuscripts at UVA
- Hydrangea
 - For articles
 - For data

On Deck:

- Preservation Repository – administrator's interface
- Digitization Workflow & Management Tool

ETDs

Search for student, advisor, title, department...

Start over | [Back to search results](#)

Actions

Submitted 2/22/2009 3:20pm

Rejected 2/22/2009 1:22:00pm
 Margin spacing is incorrect. Please reformat so margins are 1" on all sides. Then resubmit.

Resubmitted 2/25/2009 2:25pm

Next steps

- Reject (return to student to resubmit)
- Cancel (final changes in status)
- Verify (sign off each section library for final processing)

Your comment here...

#3 of 11 possible Degrees: PhD, Master, Non-Militant

[View](#) [Next](#)

Name: [Julia R. Duce](#)
 Degree: PhD
 Major: School of Earth Sciences: Interdisciplinary Program in Environment and Resources (IPER)
 Advisor: 1 advisor, 10th
 Dissertation Title: Sustainability in small-scale fisheries: An analysis of ecosystem impacts, fishing behavior, and spatial management using participatory research methods

Thesis or Dissertation (PDF)

SustainableSmallScaleFisheries.pdf	1.8MB	2/22/2009 7:41 pm
--	-------	-------------------

Supplemental files

cover.jpg	High-resolution cover image	125K	2/22/2009 2:52 pm
table_of_contents.pdf	Table of contents	2.595K	2/22/2009 7:40 pm
data.xls	Data for survey appendix B	220K	2/22/2009 2:52 pm
illustration.jpg	Full-size illustration from page 71	818K	2/22/2009 7:40 pm

Abstract

Anthropogenic impacts to ocean ecosystems have been well documented throughout the globe, highlighting the commercial fishing as a primary threat to marine life populations and habitats. Small-scale commercial fishing operations provide more than half of the world's seafood, employ over 90% of all fishermen, and take place along a vast proportion of the world's coastline. Their local, decentralized nature poses unique challenges for research, data collection, and management, particularly in developing nations. As a result, a focus on large, industrial fisheries has discounted the potential impacts of small-scale fisheries, leading to an assumption that they are generally more sustainable. My research on low-remote fishing operations in Maldives Islands, Maldives, aims to address these small-scale fisheries, and their sustainability, and in cases where significant ecological impacts are noted, to offer and compare several different conservation approaches. While engaging fishermen to participate directly in the data collection, I begin by conducting a series of studies that examine how small-scale fisheries interact with marine ecosystems. I quantitatively assess the potential ecological effects resulting from small-scale fisheries and their interactions in a marine habitat resulting from several common fishing practices, and find that the type and magnitude of

Components

- File upload (PDF & aux. files)
- Metadata entry
- Access & license settings
- Workflow & in page status panel
- Dissemination (to ILS, to preservation core, to Google)

Open Repositories 2010



July 6, 2010

SALT: Self Archiving Legacy Toolkit

SU LAIR

SALTworks

artificial intelligence SEARCH

Start Over Your Search: artificial intelligence Limited to: (no limits selected)

Logout | [Account](#)

Back to Results < Previous #6 of 2,512 Next >

Artificial Intelligence 1973

Title: Artificial Intelligence 1973
 Type: Paper Document
 Access: private

Series: Accession 1986-052
 Subseries: Correspondence
 Box: 3
 Folder: 33
 ID: dntid:yr604wr6818

This document refers to:

- Series: Accession 1986-052 (0,494)
- Box: 3 (304)
- Folder: 33 (145)

Open viewer in new window
 Download document

Components

- Gated discovery (authz for search)
- Direct edits to repository objects
- Crowd-sourced description & annotation
- Embedded page viewer

Open Repositories 2010



July 6, 2010

Manuscripts & Images at UVA

The screenshot shows the VIRGO interface with search filters for 'Photography'. The search results list items from the 'Holsinger Studio (Charlottesville, Va.)' collection, including 'Mrs. C.M. White: Holsinger Studio Collection, 1889 - 1939' and 'Dr. J.S. Greasy: Holsinger Studio Collection, 1889 - 1939'. The interface includes a search bar, filters, and a list of items with their titles and creators.

Components

- Blacklight on top of Fedora
- Complex, aggregate objects
- Asset delivery from Fedora
- Integrated dpatoka image streaming

Open Repositories 2010



July 6, 2010

Hydrangea – Open Access Articles

The screenshot shows the Hydrangea interface with search results for 'Open Access Articles'. The search results list items from the 'Hydrangea' collection, including 'Overreporting of Deaths From Coronary Heart Disease in New York City Hospitals, 2003', 'Costs and Savings Associated With Community Water Fluoridation Programs in Colorado', 'Epidemiology of Travel-associated Pandemic (H1N1) 2009 Infection in 116 Patients, Singapore', 'Health Status of Visitors and Temporary Residents, United States', 'Interventions to Promote Successful Re-Entry Among Drug-Abusing Parolees', and 'Hepatitis C for Addiction Professionals'. The interface includes a search bar, filters, and a list of items with their titles, creators, and deposition dates.

Components

- File upload
- Metadata entry
- Aggregate objects
- Access & license settings
- Authz w/ search and browse
- Tailored object views
- Dashboard
- Workflow

Open Repositories 2010



July 6, 2010

Timeline

- Currently in Y2 of a three year project
- Framework coding began in February '09
- First production application released in Nov '09
- **Hydrangea Beta** scheduled for public release this summer
- Next:
 - Distributed installations & development
 - From Beta -> 1.0
 - More heads

Open Repositories 2010



July 6, 2010

An Open Framework

Wiki: <http://wiki.duraspace.org/display/hydra>

List: hydra-tech@googlegroups.com

Code: <http://github.com/projecthydra/hydrangea>

Meet: Hydra & Blacklight Camp, Minneapolis, Oct 4-8
DLF Fall Forum, Palo Alto (California), Nov 1-3

Join us!



Gracias

Open Repositories 2010

July 6, 2010

Hydrangea – Data Curation Interface

***Whereon this **Hydra** son of war is born:
Whose dangerous eyes may well be charm'd asleep,
With grant of our most just and right desires;
And true obedience of this madness cur'd,
Stoop tamely to the foot of majesty.***

**-Shakespeare
Henry IV, Part 2**