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

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.

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

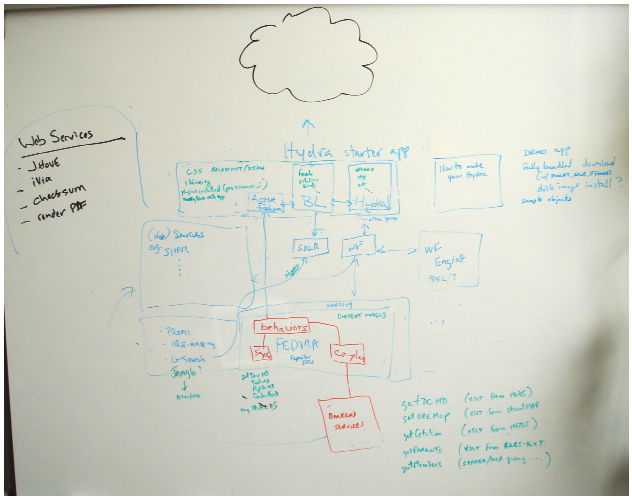
- Deposit – uploading simple or multipart objects, singly or in bulk
- Manage – editing an object's content, metadata and permissions
- Search – full text and fielded search supporting both user discovery and administration
- Browse – sequential viewing of objects by collection, attribute or ad hoc filtering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



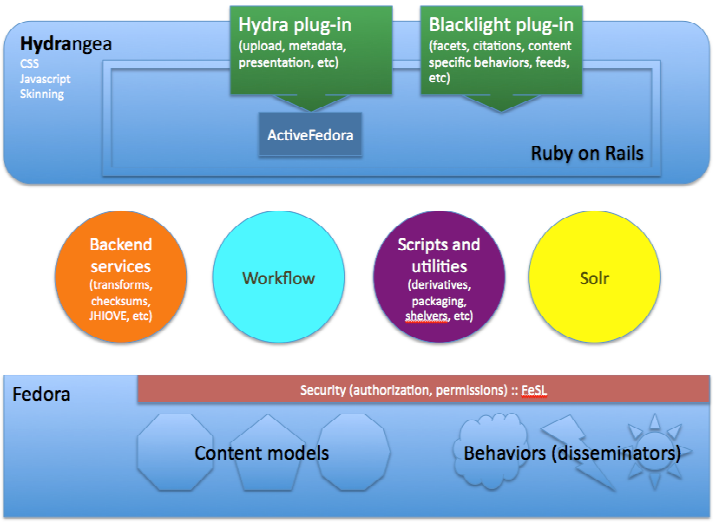
Technical Framework - Components

- **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



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 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.

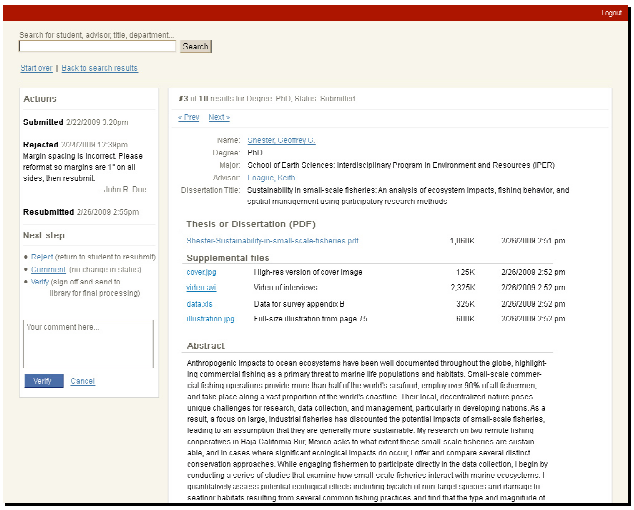
Hydra Heads

- 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

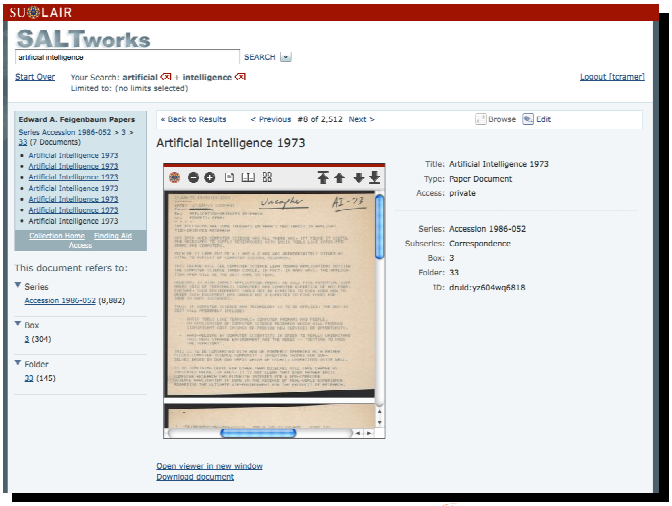
ETDs



Components

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

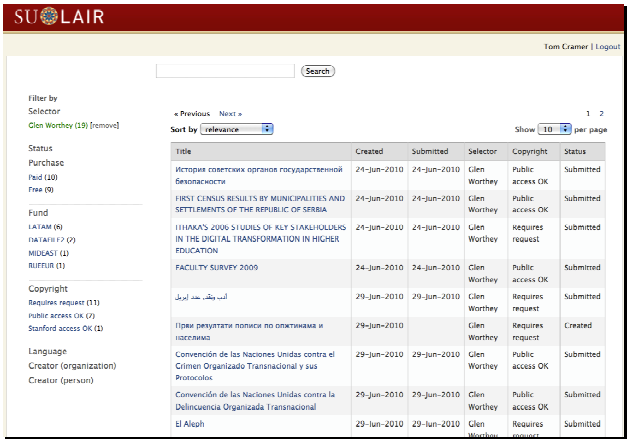
SALT: Self Archiving Legacy Toolkit



Components

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

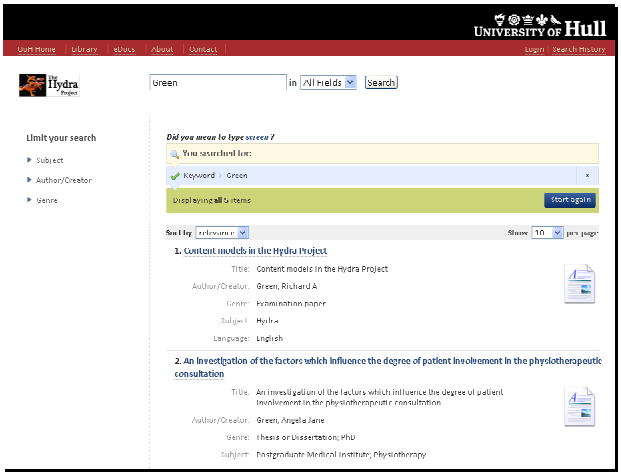
EEMs: Accessioning Born Digital Materials



Components

- File capture
- Metadata entry
- Branching workflow
- Admin "dashboard"
- Dissemination (to ILS, to preservation core, to Google)

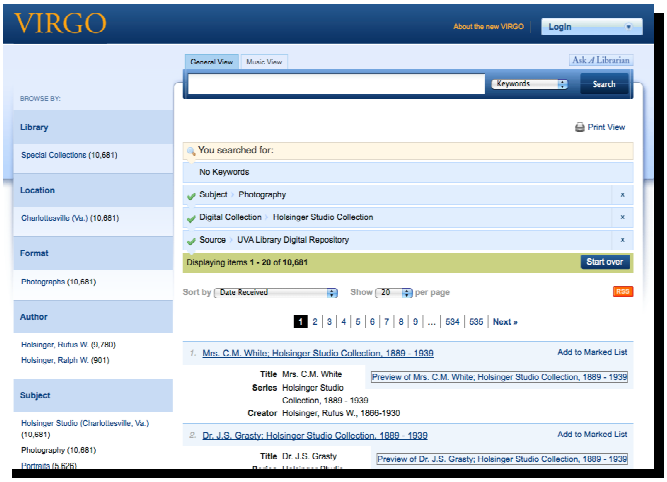
Hydra @ Hull



Components

- Blacklight on top of Fedora
- Rich, robust user search & view interface
- Configured facets to support browse
- Specialized object views
- Asset delivery from Fedora

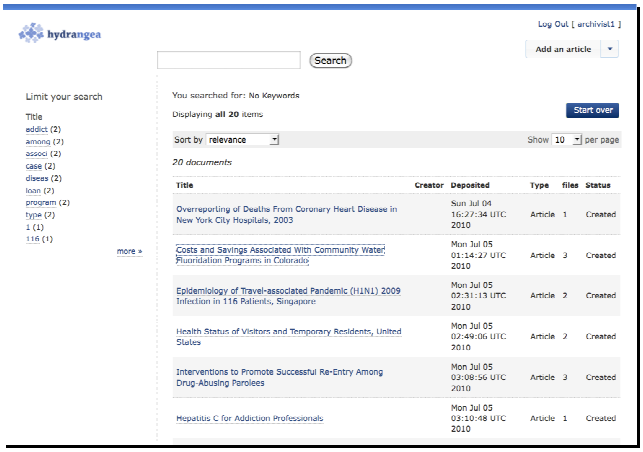
Manuscripts & Images at UVa



Components

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

Hydrangea – Open Access Articles



Components

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

Timeline

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

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

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