UNIVERSITY OF Hull

Hydra and sustainability

Chris Awre

Kultivate sustainability event, 30th September 2011

An exploration of sustainability on three levels

- Fedora/DuraSpace
- University of Hull
- Hydra

Fedora/DuraSpace

Fedora functionality

- Flexible Extensible Digital Object Repository Architecture
  - Powerful digital object model
  - Extensible metadata management
  - Expressive inter-, and intra-, object relationships
  - Web service integration (SOAP and REST)
  - Version management
  - Configurable security architecture
  - OAI-PMH conformance
  - Preservation capable
  - User interface flexibility

Background to Fedora

- Computer Science project (Sandy Payette, Carl Lagoze) at Cornell University in the late 1990s
  - Focus on how to organise digital objects
- 2002-5, Mellon funded Fedora project
  - Joint-funded project between University of Virginia and Cornell University
- 2005-7, second Mellon grant to extend Fedora development
  - Led up to release of mature production version, 2.2.1, in early 2007
- 2007-9, Fedora development moved to non-profit foundation, Fedora Commons

Fedora development

- Development has been overseen by DuraSpace since July '09
  - Parent non-profit body for Fedora, DSpace, plus Mulgara, Akubra and DuraCloud
- Fedora 3.5 now released, and clear roadmap for future development outlined
- Organisation
  - Core development team within DuraSpace
  - Community-based committers drive software development in conjunction with community input
    - Committers from US, Canada, UK, Denmark, Germany, Australia
  - Very active, and responsive, discussion lists
    - Fedora-users/Fedora-developers
  - Moving to mixed finance model: grants, sponsorship, and services
- Fedora UK&I
  - Formed 2006, meets twice-ish a year

Fedora started with a focus on

ORGANISATION

This has matured and evolved into an emphasis on

DURABILITY

University of Hull

Local need

Scalable solution (not one that has upper limit)

- Digital content is only going to grow

Standards-based (open standards where possible)

- To provide a future-proof exit strategy

Content agnosticism

- We don't know what types of content may come along

Content semantics

- Recording the relationships between different pieces of content supports future use and preservation

Areas of current activity

Datasets

Committee papers

E-prints/journal articles

Skull scan images

Student handbooks

Digitised content

Exam papers

HR documentation

Images

Open educational resources

Lectures

Audio recordings

Theses

LTSU documents

Dissertations

University policies, procedures and regulations

Developing our repository

Hull has benefitted from JISC funding through three rounds of funding

RepoMMan (2005-7), http://www.hull.ac.uk/esig/repomman/

- Focus on repository as day-to-day tool and metadata generation

REMAP (2007-9), http://www2.hull.ac.uk/discover/remap.aspx

- Focus on role of repository in supporting records management and digital preservation

CLIF (2009-11), http://www2.hull.ac.uk/discover/clif.aspx

- Focus on role of repository within digital lifecycle management across systems

Whilst these projects have addressed specific aims, implementation of an institutional digital repository has been an undercurrent through all of them

Hydra

Why Hydra?

- Our adopted interface to Fedora, Muradora, ceased development
  - Funding withdrawn
  - No community to pick it up (although some are persisting)
- Muradora was also essentially making Fedora act like a Dublin Core registry with files attached
- We wished to take fuller advantage of the richness of Fedora's capability
- We needed a development that had community as an integral part of it for sustainability
  - Acknowledging we could not go alone
- Then we presented on the REMAP project at OR2008…

Hydra

- A collaborative project between:
  - University of Hull
  - University of Virginia
  - Stanford University
  - Fedora Commons/DuraSpace
  - MediaShelf LLC
- Unfunded (in itself)
  - Activity based on identification of a common need
- Aim to work towards a reusable framework for multipurpose, multifunction, multi-institutional repository-enabled solutions
- Timeframe - 2008-11 (but now extended indefinitely)

Multipurpose, multi-institutional approach

- A repository should be an enabler, not a constraint
  - Repositories have been put forward as potential solutions for a variety of use cases
  - Hydra recognises that repositories can be used in the management of digital content at different stages in the lifecycle of that content
  - It is therefore useful to consider how to enable multiple interactions with a repository for different purposes

Hydra take-up and embedding

- Hydra is about developing flexible interfaces over a repository (in our case Fedora) that allow for the management of different types of content in the same repository
  - Hydra aims to support embedding by allowing a single repository to serve multiple needs
  - Hydra aims to support take-up through the flexible development of end user and management interfaces that are designed for use according to content type
- Hydra seeks to provide a framework to support adaptability

Fundamental Assumptions

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

…yet each needs the flexibility to tailor solutions to local demands and workflows.

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.

"If you want to go fast, go alone. If you want to go far, go together"
(African proverb)

Collaboration, collaboration, collaboration…

- Hydra is about working together in two ways
  - It has started as a collaboration between like-minded partners who have identified a shared set of issues and needs around the development of repositories
  - It recognises that the initial partners can lay out the groundwork, but that others should be able to benefit from this and build on it.
- Hydra is therefore seeking to enable
  - A community of developers and adopters extending and enhancing the core
  - Collaboratively built "solution bundles" that can be adapted and modified to suit local needs – Hydra heads!

Hydra partnerships

- From the beginning key aims have been and are:
  - to enable others to join the partnership as and when they wished (MediaShelf LLC have since joined, Northwestern, Notre Dame, Indiana, and Rock 'n' Roll Hall of Fame waiting in the wings)
  - to establish a framework for sustaining a Hydra community as much as any technical outputs that emerge
- Establishing a semi-legal basis for contribution and partnership
  - All partners sign a Memorandum of Understanding
  - All code contributors will be asked to sign a code licensing agreement
    - At institutional level
    - At individual level
  - Model based on Apache Software Foundation

Hydra content modelling guidelines

- The Hydra project has developed guidelines around the organisation and structure of content which can then be implemented using its technology stack – Hydra compliant objects
  - Exploiting the Fedora digital object model
    - Although the guidelines could also be implemented using other technologies
- Guidelines allow common and shared implementations of digital content management within Fedora, facilitating interoperability

Hydra technical implementation

- Fedora
  - All Hydra partners are Fedora users
- Solr
  - Very powerful indexing tool, as used by…
- Blacklight
  - Prior development at Virginia (and now Stanford/JHU) for OPAC
  - Adaptable to repository content
- Ruby
  - Agile development / excellent MVC / good testing tools
- Ruby gems
  - ActiveFedora, Opinionated Metadata, Solrizer (MediaShelf contributions)

Hydra community structure

- Hydra Steering Group
  - Comprises original founding members
- Hydra Partners
  - Anyone who would like to join the initiative on a formal basis
  - Must be willing to contribute back
- Hydra Developers
  - Active community of code contributors
    - See http://www.ohloh.net/p/hydra-head for analysis
- Hydra Design
  - Focus on modelling and management of content
- Hydra Adopters
  - Anyone using the models and/or code

Hydra process

- Face-to-face meetings
  - First in September 2008
  - Every 3-4 months on average since then (and ongoing)
- Bi-weekly Skype calls (watching the time difference!), alternating between Hydra Steering and Hydra Partners
- Email discussion list – Steering Group
- Google Groups – hydra-tech and hydra-users
- Wiki (part of DuraSpace wiki)

Where to learn more…

Web: http://projecthydra.org

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

List: hydra-tech@googlegroups.com / hydra-users@googlegroups.com

Code: http://github.com/projecthydra/

JIRA: https://jira.duraspace.org/browse/HYDRA

Meet: Hydra Camp, Minneapolis, 17-21 October

Next Hydra Partners meeting, Stanford, 7-9 December

Contacts and links

Chris Awre (c.awre@hull.ac.uk)
Richard Green (r.green@hull.ac.uk)
Simon Lamb (s.lamb@hull.ac.uk)

Blogs: http://hydrangeainhull.wordpress.com
http://blacklightathull.wordpress.com

Repository site: http://hydra.hull.ac.uk

Prezis (from OR11):
http://prezi.com/1lmhfhcvjhmm/hydra-technical-framework/
http://prezi.com/tf_dcoaaqguz/hydra-at-or2011/

Thank you