

Using Hydra's flexibility to manage datasets

Chris Awre, Simon Lamb, Richard Green

Components of Institutional Research Data Services

JISC MRD / DCC Institutional Engagements meeting

24-25th October 2012

An institutional repository

One institution = one repository?

- Repositories are infrastructure
 - Maintaining infrastructure requires resource, which we need to minimise to justify costs in the long-term
- Content doesn't sit in silos
 - One repository facilitates cross-fertilisation of use
- Integration with one system
 - Embedding the repository means linking to one place

Five principles

- A repository should be content agnostic
- A repository should be (open) standards-based
- A repository should be scalable
- A repository should understand how pieces of content relate to each other
- A repository should be manageable with limited resource

Five principles (leading to our implementation)

- Fedora is content agnostic
- Fedora is (open) standards-based
- Fedora is scalable
- Fedora understands how pieces of content relate to each other
- Fedora (and Hydra) is manageable with limited resource
 - With help from the community



Fedora and Hydra



- Fedora can also be complex in enabling this flexibility
- How can the richness of the Fedora system be enabled through simpler interfaces and interactions?
 - The Hydra project has endeavoured to address this, and has done so successfully
 - Not a turnkey, out of the box, solution, but a toolkit that enables powerful use of Fedora's capabilities through lightweight, Ruby-based tools
 - Principles can also be applied to other repository environments
- Hydra 'heads'
 - Single body of content, many points of access into it

Hydra is a Repository Solution



Hydra is a repository solution that is being used by [institutions on both sides of the North Atlantic](#) to provide access to their digital content. Hydra provides a versatile and feature-rich environment for end-users and repository administrators alike.



Hydra is a Community



Hydra is a large, [multi-institutional collaboration](#). The project gives like-minded institutions a mechanism to combine their individual repository development efforts into a collective solution with breadth and depth that exceeds the capacity of any single institution to create, maintain or enhance on its own. The motto of the project's partners is "if you want to go fast, go alone. If you want to go far, go together."

Hydra is a Technical Framework



Hydra is an ecosystem of components that lets institutions deploy robust and durable digital repositories (the body) supporting multiple "heads": fully-featured digital asset management applications and tailored workflows. Its principle platforms are the Fedora Commons repository software, Solr, Ruby on Rails and Blacklight. See how you can [get started](#).

Hydra is Open Source Software



Hydra software is free and open source, available under an Apache 2 license.



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.

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 as adaptable and evolving beast



- Repository activity revolves around CRUD
 - Create, Read, Update, Delete
- Repositories produce templates to enable human interaction with each of these and for internal technical management purposes
- Hydra models allow different templates for different content types, so as to more flexibly adapt to their management needs
- User interface – more sophisticated and relevant interaction
- Internal management – content is held appropriately and can be re-purposed specifically



Organise the content

-----Processed HMAP Datasets

----Law

---Scarborough School of Education

--Campus Based Publishing System (CBPP)

----cbppplatform objects

----hull-cbpp objects

--Committee Section

----General documents

--Datasets

----Domesday

-----Databases

-----Documentation

-----Phillimore appendices

-----Phillimore county introductions

-----Phillimore county notes

-----Statistics notes

-----Translation

----English

-----Repertorium of Middle English Prose Sermons

----History of Marine Animal Populations (HMAP)

-----Processed HMAP Datasets

Structural sets

Larkin Centre events

Resource type Display set

Description A collection of recordings made at public events organised by the University's Larkin Centre featuring interviews with modern writers.



History of Marine Animal Populations (HMAP) datasets

Resource type Display set

Description The HMAP datasets are a research resource comprising of information derived largely from historical records relating to fishing catches and effort ...



Skills for scientists : employability resource pack

Resource type Display set

Description This pack is a structured resource that can be used as a complete module or as individual exercises to support other career-based activities. The a...



Domesday dataset

Resource type Display set

Subject Domesday Book

Description A collection of data about and around the 1086 Domesday book. The dataset was originally lodged in the repository by Professor John Palmer in 2008....



Display sets

Work of faculty and students

Faculty

- Disseminate research outputs / learning materials
- Manage research data
- Archive event outputs

Students

- Disseminate theses / dissertations
- Provide exam papers
- Student handbook archive

Granular security required to manage these different activities

The repository has been tied into our institutional CAS system

Materials can be open, internal, or restricted to user groups / users

Work of faculty and students

Faculty

- Disseminate research outputs / learning materials
- **Manage research data**
- Archive event outputs

Students

- Disseminate theses / dissertations
- Provide exam papers
- Student handbook archive

Granular security required to manage these different activities

The repository has been tied into our institutional CAS system

Materials can be open, internal, or restricted to user groups / users

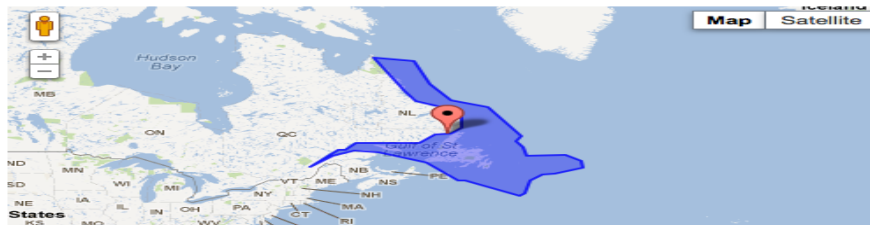
- Work to develop a data management plan for History
- Used discussions with History to adapt Hydra to better manage datasets, including Google coordinate link and citation

Dataset

HMAP Dataset 06: Newfoundland, 1675-1698

Person	Pope, P. (Author)
Subjects	Population census; History of marine animal populations; Fishing effort; Cod fishery
Rights	Creative Commons Licence: Attribution-Noncommercial-Share Alike 2.0 UK: England and Wales. See: http://creativecommons.org/licenses/by-nc-sa/2.0/uk/
Citations	(a) The dataset: please cite as follows: P. Pope, ed. 'Newfoundland, 1675- 1698' in M.G Barnard and J.H Nicholls (comp.) HMAP Data Pages (www.hull.ac.uk/hmap); (b) Supporting documentation: please cite as follows: P. Pope, 'HMAP dataset 6: Newfoundland, 1675-1698, Supporting Documentation', in M.G Barnard and J.H Nicholls (comp.) HMAP Data Pages (www.hull.ac.uk/hmap)
Description	Fishermen, settlers and cod catches in 17th-century Newfoundland. The map below gives an indication of the extent of the Newfoundland-Labrador shelf; the 'view as map' link in the download panel at the right will show a much more detailed representation. The kml file download, when used with Google Earth, will render the extent of the Newfoundland-Labrador shelf in detail.

Coverage	Newfoundland
Temporal	1675-1698
Geo-data	



Downloads

- Database - ASCII format (zip)
- Database - Access 2000 format (zip)
- Database - csv format (zip)
- Documentation - PDF format (pdf)
- Documentation - text format (txt)
- Documentation - Word (.doc) format (doc)
- Newfoundland-Labrador Shelf - kml file (99 KB vnd.google-earth.kml+xml)

QR code link to this page



DOIs and persistent identifiers

- Everything in Fedora/Hydra has a persistent unique identifier
 - This can be quoted, and forms part of the URL to the dataset (or, more specifically, the splash page)
- Interaction between Hydra and DataCite has been explored to enable the additional benefits of this widely used citable standard identifier
 - More than just an identifier

Moving on

- Staff in History now being trained to deposit their own datasets
- The work thus far has enabled us to exploit our institutional repository's flexibility to use it for datasets
 - We shall also be using it as a data catalogue
- We know we will need to exploit this flexibility further as different research data needs emerge
- Institutional, cross-Faculty, working group is being formed to oversee the implementation of the roadmap and advise on future repository developments to meet needs

EPSRC roadmap

- The EPSRC roadmap has helped structure and inform institutional priorities for research data management
 - Principles apply across all subjects, not just EPSRC areas
- Hydra is core to this
- Ongoing issue of where to actually store the data
 - Locally
 - In the cloud
 - Options are being explored
 - A mixed environment is likely

Finch outcomes

- Whatever you may think of the Finch Report (Government report on expanding access to research publications)
 - It did highlight the valuable role repositories can play in providing supporting infrastructure
- It focused on the need to have a way of managing:
 - Different types of grey literature
 - Theses/dissertations
 - **Research data** and the link to publications
 - Preservation
- Looks like an opportunity!

Hydra community



- Hydra is a community
 - We work together to identify needs and create solutions
 - Hydra's flexibility means we use the same infrastructure, but can generate individual solutions
- Hydra started with four partners
 - It now has 10 and others are looking to join in
- Hydra UK
 - Event at LSE, 22nd November – come and find out more!
 - Contact Chris Awre (c.awre@hull.ac.uk) for details
- How can JISC MRD / DCC support community initiatives like this?

Looking forward...

Our repository is still a work in progress - probably always will be

...it is maturing

We have been able to apply it to a range of purposes, including research data

It is a repository for the institution as a whole

Thank you

Chris Awre – c.awre@hull.ac.uk

Simon Lamb – s.lamb@hull.ac.uk

Richard Green – r.green@hull.ac.uk

Hydra at Hull – <http://hydra.hull.ac.uk>

Hydra Project – <http://projecthydra.org>

