

**Building the Hydra Together:
Enhancing Repository Provision through Multi-Institution Collaboration**
PAPER PROPOSAL for Open Repositories 2011, Austin, TX

Introduction

The Hydra project is a digital repository initiative started in 2008 that originally brought together three institutions (Stanford University, the University of Virginia and the University of Hull) and DuraSpace, with a common identified need to provide a flexible means for managing and delivering a wide range of digital content types. The project has since investigated and worked towards a reusable framework for multipurpose, multifunction, multi-institutional repository-enabled solutions. Two previously identified assumptions have underpinned the work:

- 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

The Hydra project has tested out these assumptions and reports in this presentation the outcomes from applying them to the work undertaken.

The original purpose of coming together was a recognition specifically of the first of these assumptions, and realising that we were likely to be better placed working together on how to address our digital content management needs rather than trying to do this alone. There was also an early acknowledgement that any output from such collaboration would only work sustainably if it could be built on and further developed by others as well. From the beginning a key aim has thus been to enable others to join the open source Hydra project as and when they wished, and to establish a framework for sustaining the community and collaboration as much as any technical outputs that may emerge.

This approach was not, however, intended to result in multiple repository solutions as, reflected in the second assumption, a common repository infrastructure was considered to be, and remains, core to sustainability. Whilst geared towards enabling a technical solution, the focus of the Hydra project has been on how a repository could be used to address the multiple use cases we identified across our institutions, and others, in a way that allowed for software implementation based on a common infrastructure to meet these different needs. At the heart of the Hydra project has been the way that content and metadata are structured within the

repository. A temptation is to be very detailed to provide a strong structure for the repository. Whilst recognising that individual repositories may wish to apply this detail, the Hydra project has adopted a simpler approach that seeks to allow different types of content to be modelled using common, atomic, building blocks. This provides the basis from which others can implement their own Hydra solutions.

Building on the work undertaken over the past three years, and having tested the assumptions through this work, this presentation will describe the results through three spheres the Hydra project is engaging with to specifically facilitate community collaboration: the underlying philosophy and governance to guide current and future community development; the aspects of the technical framework that enable other institutions to benefit and also contribute back; and the ways in which the Hydra project's approach helps to overcome common issues with repository development and sustainability.

Hydra philosophy and responsibilities

The use of the term 'Hydra' is deliberate – one body, many heads. By providing multiple points of access onto a common repository more people can interact with the repository in different ways to meet their digital content management needs. Whilst not building out the community in itself, placing the repository at the centre of digital content management infrastructure provides the impetus for sustaining its use. For the common core to work effectively it needs to be based on an open architecture that encourages contribution as use is made of it. Technical development on the Hydra project is basing its common infrastructure on Fedora, a common technical link between the founding institutions, though the principles used in developing Hydra solutions could be applied to other platforms.

The founding institutions have been equal partners throughout this development, and as Hydra Partners, now, have formed the current basis of the governance of the Hydra project. An important step through this process has been to include additional partners as appropriate: MediaShelf LLC came on board in 2010, and Notre Dame University has recently also joined the Partners (now six in number). In actively seeking to expand the group a set of roles and responsibilities has been established to sustain ongoing development that all Hydra Partners can, and will be expected to, contribute through.

There are four roles that can be undertaken as part of the Hydra project. The first three are small, coordinating bodies whilst the fourth is open to anyone with an interest in using outputs from the Hydra project:

- Hydra Steering – This role is carried out by a core of the Hydra Partners, and has the responsibility for collaborative roadmapping and resource coordination. It is also responsible for governance of the technical core, evangelism, project infrastructure, and the organisation of meetings.

- Hydra Design – This role, currently encompassing input from all current Partners, has a focus on the functionality of Hydra and how this is defined and supported. It incorporates the definition of conceptual models, design patterns, data and content models, UI design, and appropriate documentation for all. A goal for those undertaking this role is to be able to share what is produced and identify funding opportunities to support the work.
- Hydra Developer – This role is open to all with an interest in the technical approach Hydra is taking, whether Partner or not. Hydra Developers will define the technical architecture, implement the Hydra Design requirements, coordinate development through community principles, manage release cycles, and document all work undertaken for future reference. There is a regular weekly Hydra Committers call, and over 20 active contributors to the hydra-tech Google Group mailing list in the past year (all providing code, putting Hydra in the top 10% of active open source projects on Ohloh).
- Hydra Adopter – Anyone can download and run software from the Hydra project to develop their own Hydra head. Anyone can extend and modify software provided by Hydra. If you use the software, you're an adopter. There are eight heads in active development, including Partners and Adopters. They encompass solutions for open access research outputs (articles and datasets), video, images, ETDs, archival materials, and general institutional repository usage.

Getting involved

Although not the initial starting point for the Hydra project, the ultimate aim is, of course, for there to be working Hydra solutions at the partners and other institutions. This aim has informed the data modelling and technical architecture design throughout the project. Recognising local needs, Hydra Partners have adopted a pattern of designing for local need and then feeding back developments to the Hydra core reference implementation. Those wishing to implement their own Hydra solutions thus have two primary routes through which they can benefit from, and hopefully contribute back to, the Hydra community following this pattern:

- Adoption of the Hydra data model. The principle data model adopted by Hydra lays out the core metadata and structure that is required to build Hydra-compliant objects. This model can be taken by others as a baseline and used to describe their own objects, building out the model as required or using it as is. Specific implementations can be shared with others across the community: hence, use of the data model for videos at one institution can be described so that others can then benefit in their own management of videos. It is anticipated that there will in the future also be different ways in which the data model is used for the same content type, and those coming to the Hydra project should be able to select which approach suits them best.

- Adoption of software from the Hydra project. The Hydra project has produced a reference implementation, Hydrangea. This is an encapsulation of the data model to demonstrate what can be achieved. It is dynamic, insofar that ongoing work is feeding back into it, and open (the code is available through github), providing a technical starting point for others wishing to adopt the same technical stack. The founding Hydra Partners have based their implementations on the use of Ruby on Rails (soon to be using Rails 3) applications, most of which are either gems or will be soon to facilitate cross-forking and merging of code. These are layered over a Solr index of the Fedora repository. Key to enabling flexible use of this stack is the ability to deal with different content types and instances of the Hydra data model. Following a CRUD model, Create, Update and Delete are enabled through the ActiveFedora/Opinionated Metadata gems, which provide adaptive templates for different content types, and for Retrieval Blacklight flexibly displays different content types according to local requirements.

Whether these routes are followed as an Adopter or as a Partner is a local decision. We would, of course, welcome input as a Partner in order to fully realise the aim of going far by travelling together.

Why work together?

There have been many digital repository developments in recent years, many showcased through Open Repositories. Some have been adopted widely, many have not. The Hydra project has sought to learn from these experiences:

- Hydra provides a core basis upon which others can build, assured that they are developing in a way that others find useful in their own environments.
- Hydra provides a data model that can be used by others, avoiding the need to establish individual models on each occasion.
- Hydra continues to provide software that others can, and have, used to address local needs, and allowing them to focus on these needs as a priority over the underlying infrastructure.
- Hydra is implementing the governance that will allow partners to further contribute at the appropriate level or area of interest.
- Hydra is building a community of users that can provide mutual support in both development and use of the repository solutions that emerge.

Starting a repository project from scratch can be a daunting 'hill' to climb. In bringing institutions and people together, Hydra is seeking to provide a tested path over that hill.

Summary

The Hydra project is about collaboration and making the social mechanics of repositories work. This presentation will present the model we have identified and are using to further collaboration and extend the existing Hydra developments, and make our original assumptions real.