

Hydra UK discussion notes

Three breakout groups were convened at the meeting on 22nd November:

- Collection management
- Digital preservation
- Technology

Notes from these breakout groups are given below.

Collection management

This discussion started with a general theme, but quickly focused in on research data management as a key topic and issue for those present.

- RDM potentially needs complex objects
 - Hydra is currently trying to keep its approach to RDM simple
 - More complex approaches need more maintenance
 - Hydra hasn't (yet?) got detailed tooling for dealing well with children, but this is increasingly being incorporated in Hydra head developments as a requirement
- No repository will ever be able to capture all the data from an institution; it may have to fulfill a cataloguing role for stuff held elsewhere.
 - If so, is it worth using a local repository? Why not put everything into subject repositories?
 - But: there may be considerable barriers to deposit and retrieval that a local repository can better address
- Have institutions the capability to ***preserve*** data?
- What we see at the moment is the tip of a very large data iceberg
 - Academics need to be more aware of the need to manage and preserve data
 - Need to consider dealing with data from theses
 - Repositories should be capable of dealing sensibly with an original and a redacted version
- Hydra's architecture is designed to accommodate evolving needs
- Linking versions and manifestations is important

Digital preservation

Those in the group brought a range of perspectives – informed by a range of material from digitised to born-digital.

- (LSE, Oxford) - inheriting entire machines (not just files); first question "What did it look like [working]?" – emulation would be essential
- (U East London) - mix of born-digital and research data – trying to bring it all together – with preservation being a key common area
- (Northumberland Estate) - looking at repository / Sharepoint workflows and processes

Themes:

- 1) What did it look like originally?
- 2) Take what you get – preserve what you receive – versioning is critical
- 3) How to form an integral workflow?
- 4) Systems/processes often better for external depositors than internal departments – convincing people the value of preserving "stuff", that the research mandate might be useful with this.

Misc

Interesting ideas that were raised during the day include:

- The ability to capture and record additional tagging or commentary by academics or transcriptions
- Whether to ingest the disc image / tar file as 1 asset (not create an object for every file) as Oxford were, then index the tar file and use seek and sub-address
- Emulation – platforms exist for all windows OS – throw disk image at this [interesting idea but what about the broader sense of collections – ie over several accruals?] but possibly more relevant where servers had been received etc.
- How best to exchange information – i.e., other Hydra users with archival content – approaching / tackling similar issues etc.

Technology

- What constitutes a Hydra object?
 - Fedora requires DC/RELS-EXT, and Hydra requires rightsMetadata. Hydra enables complete customisation of other datastreams according to your desired object model
- How to define a content model?
 - Content models within Hydra are simply a one-to-one mapping between the RELS-EXT hasModel statement and the Ruby models you define in your Hydra application
- Hydra with other repository engines?
 - ActiveFedora, Databank, ActiveDspace? (ActiveRepository gem?)
- Interest in Ruby on Rails training opportunities
 - Omniversity in Manchester have provided a 2 day Ruby on Rails workshop
 - Suggestions, online - <http://www.edx.org/> - CS169.1x: Software as a Service – Ruby on Rails software as a service training
- Interest in a European Hydra camp

- What authentication options exist in Hydra?
 - Single users, Groups, LDAP, simple db auth? – An attendee spoke of a need to have various authentication methods in a single hydra-head.