2016-06-13 Trip Report -- Open Repositories Dublin IE

http://or2016.net

About 477 people. Mostly librarians and librarian technologists. Duraspace. Local organizing committee. Content focused on Fedora, Hydra, Dspace, and issues of digital repositories.

Monday

Hydra for Managers

Slides goo.gl/e1DhH2

 Organizing a repository to. Contain many kinds. Of things -- dissertations, books, tech reports, blueprints, images, recordings, etc.

One repository, many "heads" -- front ends for various kinds of things.

Hydra at the Boston public library.

Avalon. A time based media arrangement of the Hydra framework for Archiving film.

LMS

Black light

Repository (Fedora)

Sufia -- an arrangement of the framework. Not a head. Can be used with configuration.   Main use case is self deposit. 30 developers.

Hydra in a box will be based on Sufia. Repository product a true turn key based on Sufia, aggregator product (DPLA), hosted service (Duraspace)

Plum is a hydra head that supports digitization workflow. Developed at Princeton. Available at Github.

Oregon Digital -- linked data, originator of the Portland Common Data Model.

Hydra supports the development of. "Heads" That in turn support CRUD operations on the repository.

Active Fedora -> hydra:pcdm -> hydra works -> curation concerns -> Sufia

Hydra. "Gem" a Ruby term -- a pre-packaged something?

A hydra stack

Top. Level Sufia, Avalon, GEO data, ETDs

Curation concerns

Hydra works

Active fedora

Hydra connects these things with standard APIs. Flexible. Can handle different kinds of content.

Hydra Connect is a self-organized meeting. There is a virtual. Hydra connect.

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

Discussed with Richard Greene, Hull University. Hydra, Hydra in A Box, PCDM, VIVO.

Tuesday

Morning Plenary

Dermot Frost

Helen Shenton -- librarian, Trinity. Book of Kells, old Library. Founded by Elizabeth 1. Users of Dspace, Fedora, Hydra. Ireland's Digital economy about twice the average for a European. Nation.

William Nixon -- co-chair of OR steering committee. I have met 10 of the 18 committee members

David Minor -- UCSD. 300 submissions, 400 authors. 205 presentations/posters/workshops

?? -- idea challenge

Sarah Shreve -- U Miami. Introducing Laura Czerniewicz, University of Cape Town

Laura Czerniewicz -- Knowledge Inequalities

Cooperative Approaches to Research Data

A consortia like model for Research Data Services -- Kristi Parks, Texas Digital Library

Increasing the availability and rise of a earth Scince ontologies using an ontology

Research data management in Austria. Raman Ganguly

Duraspace is Open

Debra Hanson Kurtz, Jonathan Markow presented on Duraspace principles and practices.

Dean Krafft shared that VIVO is linked data and has experience with linked data.

Interest in working across projects.

Interest in working with international consortia such as OpenAire and Cineca.

Research Data Workflows

Hepdata.net, Eamonn Maguire. 65,000 high energy physics datasets (50K each, 10 tables of data per). All lives, exportable by JSON. Based on Invenio, a framework for publication and data management by CERN.

Data management in structural biology. Piotr Silz, Harvard. SBGrid. Integration with Protein Data Bank. Data publication, curation/analysis, data access, sustainability (preservation). National Data Services, Globus, Dataverse project. Works with Merce Crosas.

Digital Rocks. Matt Hanlon. Http://www.digitalrocksproject.org Texas advanced computing center. Volumetric images of hydro geology. Project. Samples. Origin/raw data. Analysis/derived data. Data files.

Poster Session

Graham and I presented an OpenVIVO poster. Available on figshare here: https://goo.gl/aduZXZ

Wednesday

Identify yourself. Identifiers and attribution

Giving researchers credit for their data. Neil Jeffries, Oxford. Submit "data paper". Describe data in more detail than the scientific paper. Researcher gets credit for paper, preservation of data, avoid publisher submission system. Uses dataCite and ORCiD. Submit to journal using journal template. JISC research data spring. Analysis of RDA workflow. Implemented in Fedora4/Hydra/Sufia. Positioned for RDF.

Accelerating repository workflow. Liz Krznarich, ORCiD. ORCiD 21 staff, 276 members. ORCiD is open source. 2.3 million ORCiD identifiers. 18 publishers require ORCiD. ORCiD profile is automatically updated when dataCite or cross ref doi is minted. Users must grant permission for auto update. Cross ref has supplied 800K doi with ORCiD since Oct 2015.

Integrating identifiers into national infrastructures. Michaela Minnielli. Andre Bellini. Tom Demeranville. Cineca 900 employees, founded in 1969 for high performance computing. Cineca facilitating national adoption of ORCiD in Italy. 97% of researchers authorized use of their ORCiD. Create ORCiD requires faculty opt-in and member ID. 80K of 82K Italian researchers now have an ORCiD.

Metadata and identifiers

RIOXX: a modern metadata application profile. Paul Walk. University of Edinburgh. 11 terms from Dublin core, 2 from NISO, 8 from new namespace rioxxterms. Identifiers must be http Uris. For orgs: Http://ISNI.org/isni/0000111122223333. License ref. http://Creative Commons.org/licenses/by/4.0 with a start date. [would be easy to implement in VIVO -- metadata for each paper]. DateAccepted is difficult to ascertain. Uris for licenses can be tough. Uri for a. Project (UK) also difficult. Implemented in 56 repositories in the UK. Http://www.rioxx.net

COAR Resource types -- a Skosified vocabulary for Open repositories. 100 members. COAR Controlled vocabularies for repositories. Resource types, access rights, document version types, date types. VocBench for vocabulary Editing, multilingual labels, alignment with other vocabularies. Purl.org/COAR /resource_type/c_8042 (working paper). 50 concepts. Hierarchical. Matches to other vocabularies. Reconciled Casari, dataCite, OpenAire, many others. Dspace. Prototype by university of Minho in Portugal. Eprints. Phaedra.org in Austria. OpenAire repository manager guidelines. Version 1.1 due now. Version 2 next year. http://purl.org/coar/igcv. Work with RDA vocabulary group.

Enhancing citation-only repositories. Mark Hakkarinen. Georgetown University. Stored citation only data in Dspace. Year long clean up process to upgrade 300,000 citations on bioethics collected over 40 years to modern standards. Usage went up by a factor of 4. EthxWeb.

Idea Challenge

A team including Tom Cramer of Stanford, Jiri Kuncar of CERN, Justin Simpson of Artefactual, Graham and I won the OR IDeaChallenge for a presentation entitled "The Magic Place". Slides here: https://goo.gl/zi2ZJD Basic idea is to provide a single drag and drop widget that would have multiple workflows behind it to facilitate faculty contribution of work. The widget could appear on a VIVO profile, a library repo site, OSF, or other places where faculty might be found. The faculty member would drag a file on to the magic place widget and "magic would happen" – the file would be annotated with approiate metadata from the institutional sign on system, and provides to various workflows for repo deposit, profile systems, RIM systems, CRIS systems, publication systems, security reviews or whever they file needs to go.

Comments

Much better conference in Dublin than in Indianapolis. Much more research content. Much more interest in data sharing. More interest in VIVO. Better access to thought leaders. Next year conference is in Brisbane, Australia.

Enterprise architectures for VIVO and repos and metadata providers are beginning to emerge.