IG Long tail of research data: session 1

Chairs:
Kathleen Shearer (Confederation of Open Access Repositories)
Wolfram Horstmann (Oxford University)
Najla Rettberg (OpenAIRE)

- full summary of the meeting http://www.rd-alliance.org/groups/long-tail-research-data-ig/wiki/full-summary-long-tail-research-data-ig-meeting.html

Long tail data = non big data (small, diverse data sets)
- managed by many diverse players in many different contexts: universities, libraries, research centers, government

- presentations http://rd-alliance.org/groups/long-tail-research-data-ig/wiki/profiles-managing-long-tail-presentations-dublin-meeting-2014

EXTERNAL SOLUTIONS

The Dryad Digital Repository
Chris Taylor
- relatively young, started 2009
- integration with journals
- curation via user support
archiving via CLOCKSS
download statistics
journal policy support
metadata syndication / APIs for discovery and retrieval

Emphasis on long tail of research data – generalist

Connect research data to ecosystem of scholarly communication standards .
(metadata obviously critical)

Scientific Data
http://www.nature.com/scientificdata/

Content type: Data Descriptor with 2 components

Benefits:
Get credit for Sharing Your Data
Open-access
Discover and Reuse
Peer-reviewed and Curated

F1000research
- presenter Varsha Khodiyar
http://f1000research.com/

- remove the publication delay
- invited peer review
- transparent refereeing
- including all data
- no restriction of access
- all article types published

Internal pre-publication checks:
- storage
- format
- layout
- labeling
- adequate data
- adequate protocol information

If no existing repository or suitable alternative, they work with figshare.
Data is viewable without leaving the article.
Data citation and cross-linking

Ubiquity Press
- Brian Hole
http://www.ubiquitypress.com/

Multidisciplinary interoperability
- 30 USD for article

Zenodo
- presenter Tim Smith, CERN/IT
http://zenodo.org/

- accept all data types, any format, any content, any licenses
- size less than 2GB through public interface
- no acquisition, no self service
- co-working with publishers, institutions
- size in TB is OK, similar to figshare

- linking with Dropbox
Export – BibTeX, DataCite, DC, EndNote, NLM, RefWorks, MARCXML
Authentication: OpenAIRE, ORCID, Google, Twitter

Zenodo connector for github - just connect to github and auto-preserve…

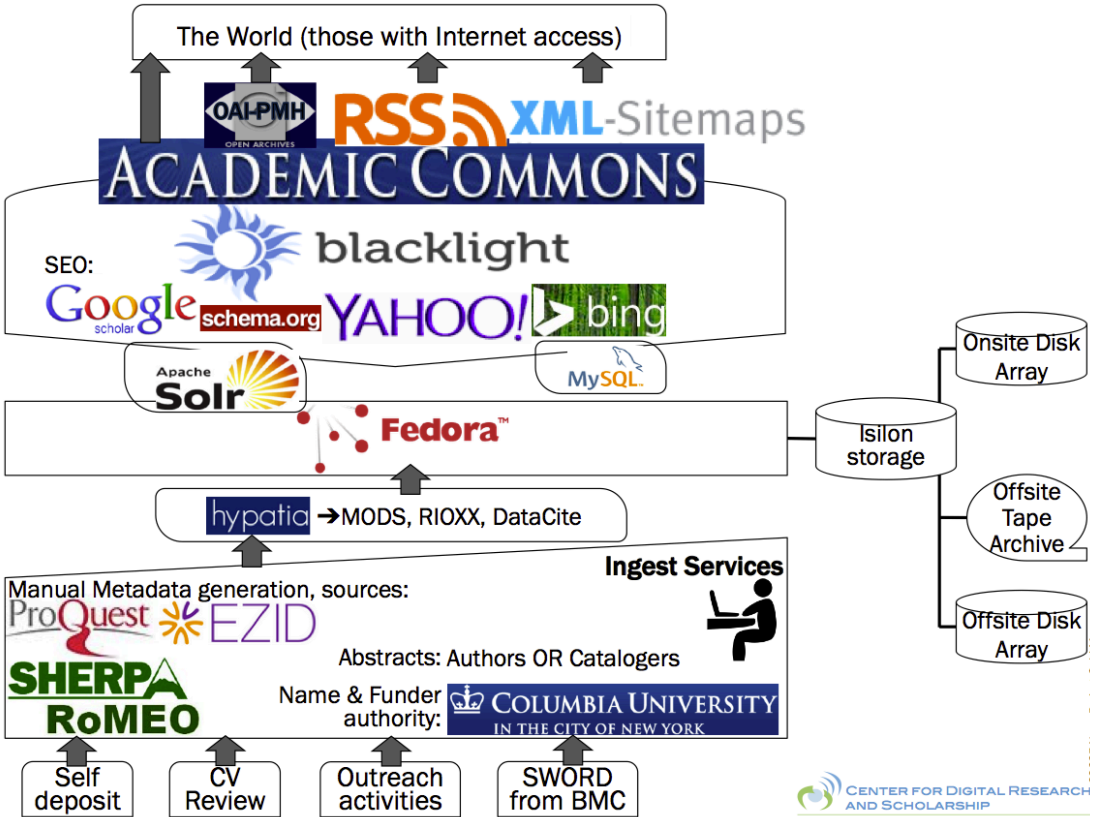
INSTITUTIONAL PROFILES

CDL Core Services
http://www.cdlib.org/services/uc3/

DMP Tool – Web app to create data management plans
Merritt – Content-agnostic repository
DataShare – data curation platform for researchers
DataUp – Web App to describe and share tabular data
EZID – service to generate and track identifiers
Web Archiving Service – archive, search websites

Columbia University ACADEMIC COMMONS
- Amy Nurnberger, Columbia University Libraries/Info Services presenting
http://academiccommons.columbia.edu

- running also archiving systems



Robin Ruggaber, **University of Virginia** discusses Hydramata project:
https://wiki.duraspace.org/display/hydra/Hydramata+Project

- sharable modules for institutional research repositories

Long Tail of Research Data: institutional services profiles
Andrew Burnham, University of Leicester

University of Leicester

Reaching out

- "Research Liaison" roles in IT and Library
- RDM and IT support for any researcher before/after bid submission
- PhD Training sessions – planning your research
- Bid costing as a trigger for guidance & support (take-up <10% ...)
- Encouragement to use DMP Online (customisation to come)
- Taking on battles with IT Services for researchers (finding solutions vs. "*we don't do that*")
- Research File Storage (and other IT Services) for all (awareness raising)
- Preservation: Dspace + iRods for Leicester Research Archive?

RDM Support & Guidance – "*What would you do if you lost your research data tomorrow?*"
- Starting point – embedding RDM into Research Computing Management Group (all Colleges represented) to direct progress
- JISC Transformations project (£, collaboration, dedication of time)
- RDM Working Group to co-ordinate and drive (Co-ordination of IT Services, Academic Practice Unit, Library, Research Support Office)
- Substantial RDM Website (as a basis for guidance)
- Data Planning documents (9 documents)
- RDM brochure - RDM Brochure
- Video

University of Leicester

Challenges

- End-to-end data management and storage
- IT Systems vs. services
- Resourcing – people, change, archive/repository...
- Accepting collaboration as the default way of working
- Responsibility, communication, & reach – "the centre", Colleges, Departments, Groups, individuals
- Grasping Open Access
- Satisfying the many and diverse vs. the few (IT literate, resourced, understand requirements and what they need)

- review of tools and services: http://www2.le.ac.uk/services/research-data

---
ORA-Data (previously named DATABANK)
- presenter Sally Rumsey
https://databank.ora.ox.ac.uk/

Oxford - putting Hydra on top of ORA-Data to improve interface
http://projecthydra.org

RESEARCH INITIATIVES

Overview of **Strasbourg Astronomical Data Centre**- three services, established 1972 - http://cdsweb.u-strasbg.fr

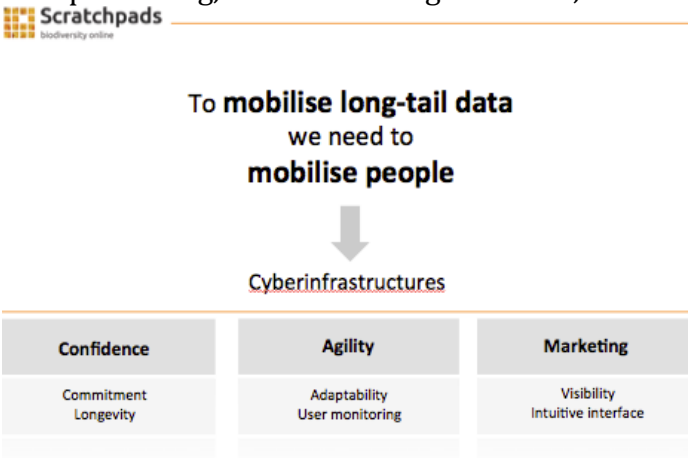
- 3 services:
    - SiMBAD – astronomical database with 7.5M objects
    - VizieR – for data
    - Aladin – interactive sky atlas

SiDORA
Thorny Staples, Smithsonian Institution presenting SiDORA built on Islandora and Fedora http://www.fedora-commons.org/profile/5

Schratchpads (biodiversity online)
- making small data… big

scratchpads.eu
- developed and maintained for last 7 years by Natural History Museum in London

- intend to enable seamless workflow for
data publishing, collection and generation, curation, analysis



IG Long tail of research data: session 2

Relationships with other WG/IS:
- Publishing Workflows
- Bibliometrics
- Publishing Data Services

Metadata Interest Group

Presentation - results of survey of current practices for discovery of research data in repositories.
Survey (February/March 2014, 60 responses, 30 responses complete)
Purpose: to what are current practices in terms of discovery metadata
Scope: any repository collecting long tail data
Platform: DSpace – 9, Fedora – 3, Eprints – 2, Islandora – 2 …

Schema: Dublin Core – 9, DataCite – 3…

DOI´s are used by less than 50%

Discussion:
1/ Are the current practices for metadata sufficient for discovery?
2/ What discoveries can we employ to improve discovery of datasets?
3/ Can we identify some good practice?
4/ What other issues could we pursue through this IG?

- make sure about machine readable metadata – sufficient descriptions

Comment by a researcher - "**Researchers needs to have training!**"