Scoping and Developing
Institutional Data Services: the
Data Libraries of 2020

IASSIST Conference
Friday 29 May 2009

Luis Martinez Uribe
Luis.Martinez-Uribe@oerc.ox.ac.uk
Digital Repositories Research Coordinator
Summary

• Data libraries
• Drivers in Oxford for the start of data related initiatives
• Projects
  – Scoping Digital Repository Services in Research
  – Embedding Institutional Data Curation Services in Research
Data Libraries

A data library refers to both the content and the services that foster use of collections of numeric, audio-visual, textual or geospatial data sets for secondary use in research.

- Long-standing tradition
- Seeds of institutional data services
Drivers in the University of Oxford

- Wealth of existing digital resources in Oxford and increasing digital content generated in the University
- Maturity of digital repository services and need of integration between data and traditional publication
- Data policies from funding agencies
Oxford Digital Repositories Steering Group

- ORA
  - Oxford Research Archive
- Oxford Digital Library
- OeRC OxGrid
- ASUC digital collections
- Dept/Faculty digital collections
- Google Materials
- Other Digital collections

- Publications & Grey Literature
  - EJ
  - Pre-prints
  - Post-prints
  - Book Sections
  - Conference papers
  - Conference presentations
  - Working papers
  - Other grey literature
- eTheses
- Grid Datasets Records
  - Referenced from OeRC
  - Technical metadata
  - Automated ingest direct to Fedora
- Other collections

Range of digital collection activities, for example:
- OxCLIC Digital Image Management Project:
  - http://wiki.oucs.ox.ac.uk/ltpg/public/OxCLIC
- MSD Research Discovery System:
  - http://research.mercury.ox.ac.uk/
- Ask Accessing and Storing Knowledge Project:
  - http://ask.oucs.ox.ac.uk/ask/index.php/Main_Page
Scoping Digital Repository Services for Research Data Management

Funded by
the OUP John Fell Fund
and the JISC funded

From Jan’08 to March’09

Aiming to scope the requirements for digital repository service to store, disseminate and preserve research data generated at Oxford.
I couldn't make sense of the data I collected for my PhD 5 years ago.

I want to publish the data as an additional resource for readers of my published book/article.

When researchers leave the department we loose all the data they created.

We collected data as part of an international collaboration but we don't know who owns the data.

Interviews with researchers

• Around 40 interviews
• Part of the Oxford case study work for the UKRDS feasibility study
• Following scenarios:
Top requirements for services

• Advice and support across research lifecycle
• Secure storage for large datasets with access controls
• Sustainable infrastructure for publication and long-term preservation
Consultation with service units

Aiming to

- Validate the researchers’ requirements for services
- Determine the data management services available to researchers in Oxford
- Identify gaps in service provision

Central University Research Ethics Committee (CUREC)
Research data management and curation services framework

| DATA MANAGEMENT & SHARING PLANS | LEGAL & ETHICAL | BEST FORMATS & BEST PRACTICE | SECURE STORAGE | METADATA | ACCESS & DISCOVERY | COMPUTATION ANALYSIS & VISUALIZATION | RESTRICTED SHARING | DATA CLEANING | PUBLICATION | ASSESS VALUE | PRESERVATION | ADD VALUE |
|---------------------------------|-----------------|------------------------------|----------------|----------|-------------------|--------------------------------------|-------------------|--------------|-------------|-------------|-------------|-----------|---------|

**TOOLS & INFRASTRUCTURE**

**ADVICE & SUPPORT**
Findings of consultation with service units

- Widespread expertise in data management and curation amongst service units in Oxford
- Support provided in ad-hoc basis but services not made explicit
- Overall, the majority of the services in the data management and curation framework are not offered fully or at all.
- There is a need for a university wide policy on data management and curation
The Data Audit Framework (DAF) in Oxford

- Funded by DISC-UK DataShare to pilot the DAF and share the experiences with the rest of the community.
- Used the DAF methodology to:
  - document two research group’s data management practices;
  - map research data resources within the two research groups and
  - compile information about data resources freely available on the Oxford University web pages.
Embedding Institutional Data Curation Services in Research (EIDCSR)

Funded by JISC
From April 2009 to September 2010

Cardio Mechano-Electric Feedback Group
Computational Biology Group

Aim is to address the data management and curation requirements of two collaborating research groups in Oxford.
The EIDCSR research data challenge

"expects research data generated as part of BBSRC support to be made available...data should be retain for a period of 10 years after completion of the project"
A lifecycle approach

1. Policies and cost models help research groups to plan and cost data curation activities

2. Data are created/received and stored on departmental local storage

3. Provenance metadata is automatically captured and stored with the data

4. Appropriate preservation metadata is added

5. Data and metadata can be stored on central data archival resources for long-term curation

6. The data is made accessible to be shared between departments and collaborators

7. The data generated by computational resources can be moved back to departmental data storage

Central data archival resources

Metadata services

Ingest

Preservation action

Transform

Create or receive

Appraise & select

Access, use & reuse

Store

Conceptualise

Computational resources

Access
Next steps

• JISC EIDCSR
  – Use the DAF to capture requirements iteratively
  – Requirements will feed into technical requirements, policy and business models
  – Identify metadata standards to describe, manage and preserve the research data

• UKRDS Pathfinder activities

• Keeping Research Data Safe 2

• To follow developments  http://eidcsr.blogspot.com
Thanks