# DISC-UK DataShare project

# *Edinburgh DataShare*: Tackling research data in a DSpace institutional repository

**Robin Rice**

EDINA and Data Library, Information Services
University of Edinburgh, Scotland

**DSpace User Group Meeting**
Gothenburg, Sweden          14-16 October, 2009

# Storyboard

- About EDINA & Data Library at UoE
- About the DISC-UK DataShare project
- What's different about data?
- Enter the Data Audit Framework
- Lessons learned from both projects

# EDINA    JISC

- EDINA is the JISC national academic data centre based at the University of Edinburgh*. Our mission and purpose is to 'enhance the productivity of research, learning and teaching' across all universities, research institutes and colleges in the UK.

- We do this by delivering first-rate online services and by working with support staff in university and colleges and with other partners … and by carrying out successful R&D projects.

# Data Library: History

- Established out of the *Program Library Unit* in early 1980s to provide access to data on mainframes, e.g. 1981 population census data.

- Part of long tradition of sharing machine-readable data for secondary analysis in the social sciences

- Formed the EDINA national data centre in 1996 - data library continues University remit
  - □ celebrated 25th anniversary in 2008

# What is a data library?
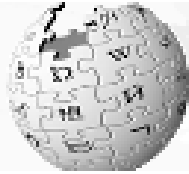


A **data library** refers to both the content and the services that foster use of collections of numeric, audio-visual, textual or geospatial data sets for secondary use in research.

A data library is normally part of a larger institution (academic, corporate, scientific, medical, governmental, etc.) established to serve the data users of that organisation. The data library tends to house local data collections and provides access to them through various means.

# Edinburgh Data Library services … distilled

- ## Finding…

"I need to analyse some data for a project, but all I can find are published papers with tables and graphs, not the original data source."

- ## Accessing …

"I've found the data I need, but I'm not sure how to gain access to it."

- ## Using …

"I've got the data I need, but I'm having problems analysing it in my chosen software."

- ## Managing …

"I have collected my own data and I'd like to document and preserve it and make it available to others."

# DISC-UK

## Data Information Specialists Committee - UK

- A forum for data professionals working in UK Higher Education who specialise in supporting staff and students in the use of numeric and geo-spatial data.

- DISC-UK's aims are -
  - Foster understanding between data users and providers
  - Raise awareness of the value of data support in Universities
  - Share information and resources among local data support staff

- We are also members of **IASSIST**, an international organisation of professionals working in and with information technology and data services to support research and teaching in the social sciences.

DISC-UK has completed a JISC-funded repository enhancement project (March 07 - March 09) with the aim of "exploring new pathways to assist academics wishing to share their data over the Internet".

With three institutions taking part – the Universities of Edinburgh, Oxford and Southampton – a range of institutional data repositories and related services have been established.

The project was led by the JISC-funded national data centre, EDINA, at the University of Edinburgh, which also runs the University's Data Library service.

# DISC-UK DataShare Partners' Repository Experience (March, 2007)

# Project Keywords



Capacity building
Best practice guidelines
Collaboration
Curation tools
Data archiving
Data Documentation Initiative (DDI)
Data Mashups
Data professionals
Data formats
Data publishing
Digital preservation
DSpace
Data sharing
EPrints.org
Funders' mandates
Edinburgh
e-Research
Fedora
Helping researchers
Institutional Repositories
Librarians
London School of Economics
Metadata standards
Managing research data
Open Access
Open Source Software
Open Data
Quantitative datasets
Orphaned datasets
Oxford
Research lifecycle
Skills & training
Secondary analysis
Southampton

"Live" cloud tag at http://www.disc-uk.org/collective.html
based on social bookmarks

# Project Briefing Papers

- Gibbs, H. (2007). **DISC-UK DataShare: State-of-the-Art Review**

- Martinez, L. (2008). **The Data Documentation Initiative (DDI) and Institutional Repositories**

- Macdonald, S. (2008). **Data Visualisation Tools: Part 1 - Numeric Data in a Web 2.0 Environment**; **Part 2 - Spatial Data in a Web 2.0 Environment and Beyond**

- Green, A., et al (2009). **Policy-making for Research Data in Repositories: A Guide**

# What's *different* about *data* ?

- **Research data are collected, not authored.**
  - ☐ Data may be shared, but are they published?
  - ☐ In a data repository, is the repository the publisher?
  - ☐ There are no explicit rewards for sharing data.
- **Size, type, complexity, update frequency**
  - ☐ DSpace is improvement on informal sharing methods.
  - ☐ Other solutions may work better for intensive data curation (see our ***Data Sharing Continuum***)
- **Who 'owns' the data? Who is the rights-holder?**
  - ☐ (individual/dept/institution/funder/subjects/nobody?)
    - ■ but minimal IPR exist in data. Issues about licensing.
- **Is Dublin Core sufficient?**
  - ☐ *Edinburgh DataShare* has set up a Dublin Core metadata schema for datasets

# *Edinburgh DataShare* Dublin Core-compliant metadata fields

Depositor (contributor)

Data Creator

Title

Alternative Title

Dataset Description (abstract)

Type

Subject Classification (JACS)

Subject Keywords

Funder (contributor)

Data Publisher

Spatial Coverage

Time Period (temporal coverage)

Language

Source

Dataset Description (TOC)

Relation (Is Version Of)

Supercedes

Relation (Is Referenced By)

Rights

Date Accessioned

# Partnerships in the Data & Research Lifecycle

**Data preservation, dissemination & long term stewardship:** *Repositories and data archives provide preservation services such as format migration and media refreshment; dataset may survive a period of dis-interest before being re-discovered.*

**Discovery and Planning** →

*Data creation, collection, repurposing:* Partnerships between researchers & support services with subject expertise; informed by domain standards and guidelines relating to formats, metadata, version control, etc.

↑ **Long term access**

**Data Analysis** ↓

**Repositories**

**PARTNERSHIPS**

**Curation services**

**Researchers**

**Data sharing and distribution:** *Repositories ingest and manage research outputs; offer federated searching, redundant storage, access controls; scholarly publications linked to data.*

← **Publication and Sharing**

*Ann Green, Digital Lifecycle Computing*

*Data processing, management and curation:* Data are transformed, cleaned, derived as part of the research process; curators identify 'partnering moments' to capture content for documentation and description. Staging repositories offer curatorial workspaces.
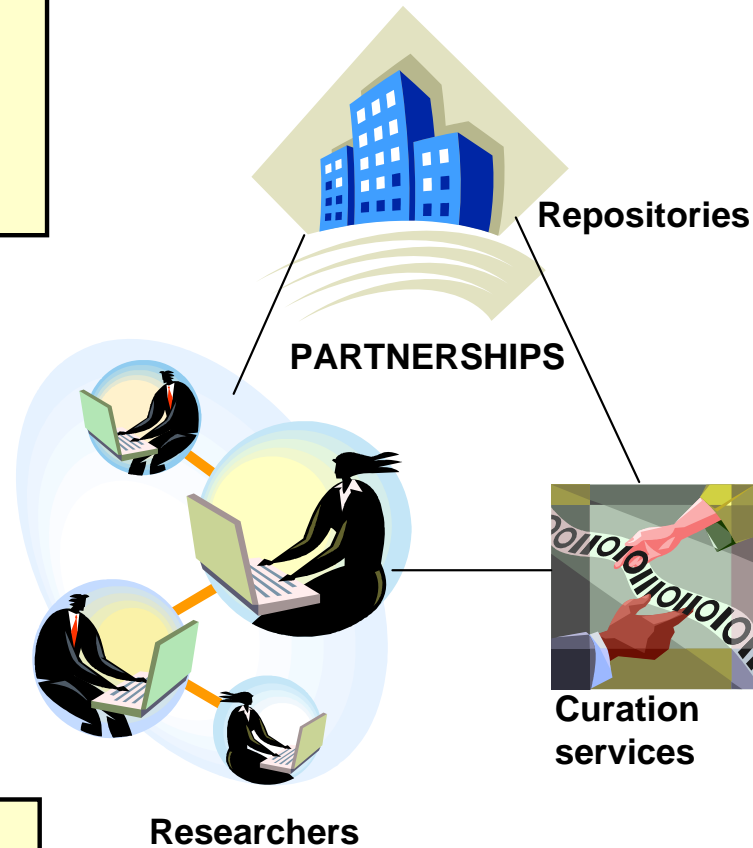
# Enter Data Audit Framework

*Recommendation to JISC:*

"JISC should develop a Data Audit Framework to enable all universities and colleges to carry out an audit of departmental data collections, awareness, policies and practice for data curation and preservation."

Liz Lyon (2007). *Dealing with Data: Roles, Rights, Responsibilities and Relationships*

# Data Audit Framework (DAF) Projects 2008

- **JISC funded five six-month projects:**
  - DAF Development (DAFD) Project, led by Seamus Ross (Director), Sarah Jones (Project Manager) HATII/DCC, University of Glasgow
  - Four pilot implementation projects:
    - King's College London
    - *University of Edinburgh*
    - University College London
    - Imperial College London
- **Two more conducted by DataShare partners, the Universities of Oxford and Southampton, as added deliverables**

Data Audit Framework

- See www.data-audit.eu
- DAF project reports available (findings)
- Appendices with questionnaires, interview schedules, etc
- Methodology document
- Online tool ready for others to conduct data audits

# Methodology

Based on Records Management Audit
methodology. Five stages:

- Planning the audit;
- Identifying data assets;
- Classifying and appraising data assets;
- Assessing the management of data assets;
- Reporting findings and recommending change.

# Lessons Learned Overall (1)

- Top-down drivers are important for overcoming barriers to data sharing (e.g. funders' requirements for data mgmt and sharing plans) as they are for open access publishing.

- Data management motivation is a better bottom-up driver for researchers than data sharing but is not sufficient to create culture change.

- Institutional repositories can play a part in overall infrastructure for data sharing

- Data librarians, data managers and data scientists can help bridge communication between repository managers & researchers (see Data Skills/Career study, Swan & Sheridan 2008).

# Swan, Sheridan 2008 …

The report calls for a 'repositioning' of the role of the library in data-intensive research. The authors of the report Alma Swan and Sheridan Brown write: 'We see three main potential roles for the library...Increasing data-awareness amongst researchers; providing archiving and data preservation services through institutional repositories; and developing a new professional strand of practice in the form of data librarianship.'

# Lessons Learned Overall (2)

- Institutions should consider developing research data policy, to clarify rights & responsibilities.
- Institutions create a broad range of data in the course of research, not just numeric datasets. So for *institutional* data repositories, the self-archiving model is probably the best for ensuring data quality. Nevertheless, researchers need guidance.
- IRs **can** improve impact of sharing data over the internet (permanent identifiers, citations, links with publications, discoverable metadata, long-term access and stewardship).
- Don't conduct institutional data audits unless you're prepared to open a can of data management worms!

# Finally

- *And don't go it alone. Get buy-in from other institutional stake-holders (computing staff, librarians, department heads, principal investigators, records managers, archivists, research office staff). Collaborate. Have fun* ☺

www.disc-uk.org/datashare.html

R.Rice@ed.ac.uk