

# Data Requirements and Digital Repositories

**IASSIST Workshop**

Tampere, Finland

26 May, 2009



# Overview

- Welcome & Introductions
- DISC-UK DataShare project background
- *About Policy-making for Research Data in Repositories - A Guide.*
- Chapter One: Content Coverage
- Facilitated break-out groups
- Break
- Report back, Discussion & Close



# Getting to know one another

## ■ One-minute introductions

- ☐ Name,
- ☐ Affiliation
- ☐ Why interested in workshop



# 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.





# Envisaged outcomes

- Exemplars of setting up institutional data repository services at each partner institution
- Enhancements to partners' IRs - with documentation and open source code for adapting DSpace, Fedora and EPrints repository software for handling datasets
- Toolkits, briefing papers and other outputs to inform UKHE repository community about data management and research support
- Technical watch on e-Research, VREs, Web 2.0 and related developments
- Papers, presentations and online dissemination of collected knowledge

# Blogging and bookmarking



Actual live cloud tag is on project's *Collective Intelligence* web page based on social bookmarks collected on *Faves*.



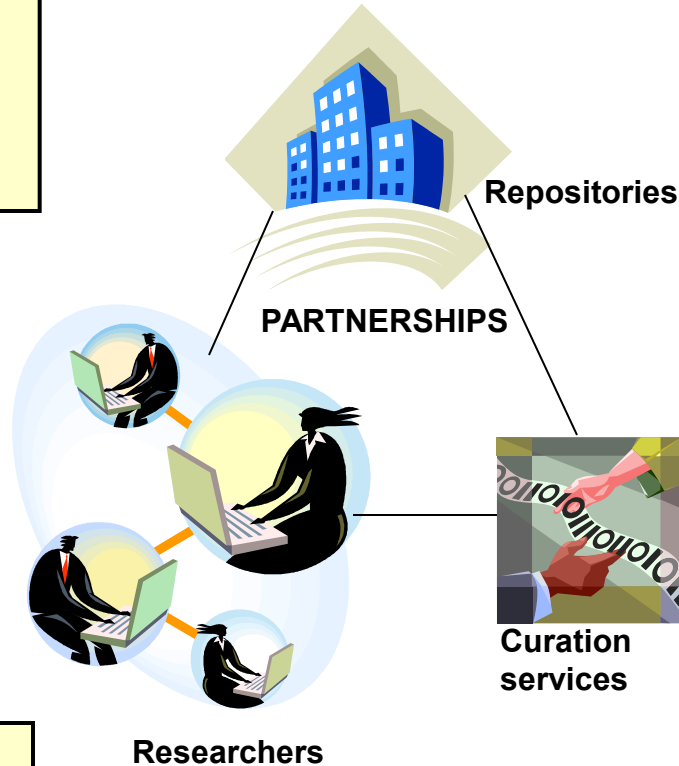
**Data preservation, dissemination & long term stewardship:**

*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.*

Long term access

**Partnerships in the Data & Research Lifecycle**

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.*

Data Analysis

**Data sharing and distribution:**

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

**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.*

Publication and Sharing

*Ann Green, Digital Lifecycle Computing*



# 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*



- See [www.data-audit.eu](http://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

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

# 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 rectangular datasets. So for *institutional* data repositories, the self-archiving model is probably the best for ensuring data quality. (Repository is a host, not a publisher. Only metadata is moderated.)
- 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!



# Goals of the Guide

- A tool for working through the requirements for taking data into digital repositories
- Highlight issues to consider when IRs start taking in data
- Optimize experience from both communities: digital repository managers and data managers/librarians
- Facilitate cross-talk and the development of strategies to deal with data
- A first step toward making policies and building system requirements (or considering new features like new content models or taking in new metadata schemas) and depositor agreements
- A compilation of DATA related requirements from multiple sources that didn't focus upon DATA



# Sources used to create the Guide

Reviewed and incorporated pieces of:

- Assessment of UKDA and TNA Compliance with OAIS and METS standards. Questions in the Appendix got us started.
- OAIS reference model for concepts and definitions
- Liz Lyon: Dealing with Data
- RIN: Stewardship of Digital Research Data
- Oxford University Research Archive policies
- OpenDOAR policy making tool, the DataShare participants used the tool to draft policies or compared them to their existing policies.
- ICPSR and UKDA guidelines for preparing data for submission and other documents.
- TRAC checklist, DRAMBORA, and nestor: repository requirements and assessment tools.
- a range of DATA QUALITY resources, notably DANS