

Executive Summary

The DISC-UK DataShare Project was funded from March 2007-March 2009 as part of JISC's Repositories and Preservation programme, Repositories Enhancement strand. It was led by EDINA and Edinburgh University Data Library in partnership with the University of Oxford and the University of Southampton. The project built on the existing informal collaboration of UK data librarians and data managers who formed DISC-UK (Data Information Specialists Committee – UK).

This project has brought together the distinct communities of data support staff in universities and institutional repository managers in order to bridge gaps and exploit the expertise of both to advance the current provision of repository services for accommodating datasets, and thus to explore new pathways to assist academics at our institutions who wish to share their data over the Internet.

The project's overall aim was to contribute to new models, workflows and tools for academic data sharing within a complex and dynamic information environment which includes increased emphasis on stewardship of institutional knowledge assets of all types; new technologies to enhance e-Research; new research council policies and mandates; and the growth of the Open Access / Open Data movement.

With three institutions taking part plus the London School of Economics as an associate partner, a range of exemplars have emerged from the establishment of institutional data repositories and related services. Part of the variety in the exemplars is a result of the different repository platforms used by the three project partners: DSpace (Edinburgh DataShare), ePrints (e-Prints Soton) and Fedora (Oxford University Research Archive, ORA)--all open source software. LSE took another route and is using the distributed Dataverse repository network for data, linking to publications in LSE Research Online. Also, different approaches were taken in setting up the repositories. All three institutions had an existing, well-used institutional repository, but two chose to incorporate datasets within the same system as the publications, and one (Edinburgh DataShare) was a paired repository exclusively for datasets, designed to interoperate with the publications repository (Edinburgh Research Archive).

The approach took a major turn midway through the project when an apparent solution to the problem of lack of voluntary deposits arose, in the form of the advent of the Data Audit Framework. Edinburgh participated as a partner in the DAF Development project which created the methodology for the framework, and also won a bid to carry out its own DAF Implementation project. Later, the other two partners conducted their own versions of the data audit framework under the auspices of the DataShare project.

A number of scoping activities were carried about by the partners with the goal of informing repository enhancement as well as broader dissemination. These included a State-of-the-Art-Review to determine what had been learned by previous repository projects in the UK that had forayed into the data arena. This resulted in a list of benefits and barriers to deposit of datasets by researchers to inform our outreach activities. A Data Sharing Continuum diagram was developed to illustrate where the projects were aiming to fit into the curation landscape, and the range of curation steps that could be taken, from simple backup to online visualization. Later on, a specialized metadata schema was explored (Data Documentation Initiative or DDI) in terms of how it might be incorporated into repository systems, though repository development in this area was not taken up. Instead, a dataset application profile was developed based on qualified Dublin Core (dcterms). This was implemented in the Edinburgh DataShare repository and adapted by Southampton for their next release. The project wished to explore wider issues with open data and web publishing, and therefore produced two briefing papers to do with data mashups – on numeric data and geospatial data. Finally, the project staff and consultant distilled what it had learned in terms of policy development for data repositories in a training guide. A number of peer reviewed posters, papers, and articles were written by DISC-UK members about various aspects of the project during the period.

Key conclusions were that 1) Data management motivation is a better bottom-up driver for researchers than data sharing but is not sufficient to create culture change, 2) Data librarians, data managers and data scientists can help bridge communication between repository managers & researchers, and 3) IRs can improve impact of sharing data over the internet.