



## Progress Report

<b>Project Name</b>	DISC-UK DataShare (EDINA)
<b>Project Website</b>	<a href="http://www.disc-uk.org/datashare.html">http://www.disc-uk.org/datashare.html</a>
<b>Report compiled by</b>	Robin Rice
<b>Reporting period</b>	Nov 2007 – Apr 2008
<b>Section One: Summary</b>	
<p>DataShare has been active at a time when research data is becoming a focus for the repository community, and the project team is committed to delivering on the important deliverables that will act as exemplars for others to learn about this tricky area before diving in. Despite an initial commonality around social science data, it has become apparent to most of the partners that it will need to devise strategies to handle a wider range of data types and support depositors across their institutions. This in turn has shown the importance of joined-up working within our institutions, and the need for advocacy not only among the research community, but with other information service staff, such as those providing storage solutions, collaborative software tools, policy makers, and domain specialists, wherever they are based in the University.</p> <p>Transparency and dissemination have been emphasised throughout this short project. As soon as a deliverable used by the team to look at metadata options (DDI and Dublin Core) was ready, it was simultaneously published on our website and announced to relevant lists, for example. We have used our blog to report on events, and the Collective Intelligence page of our website to pipe in other blogs reporting on other relevant topics. We have continued to use a social bookmarking tool to generate on-the-fly bibliographies, and set up a dynamic tag cloud on our website to show the topics collected there. All of us have had a steep learning curve, and continue to benefit from having experienced data support professionals as well as experienced repository managers and developers among us.</p>	
<b>Section Two: Activities and Progress</b>	
<p><i>WP 1: Project Management</i></p> <p>Regular communication was maintained through the project list and through telecon meetings. A face-to-face meeting at EDINA in February brought project partners, the consultant, the programme manager, and local experts, including Charlotte Waelde from the AHRC Centre for Intellectual Property Law, together for two days of intensive discussion and talks. This helped with decision-making for each site on WP 2. Opportunities for professional development were raised to project staff; reporting or other outputs from events were requested. The website was maintained as was communication with JISC throughout.</p> <p><i>WP2: Institutional Repository Development</i></p> <p>Please see individual progress reports.</p> <p><i>WP3: Technology</i></p> <p>Most advanced workpackages have not started yet, until issues from WP2 are resolved.</p> <p><i>WP4: Outreach and Dissemination</i></p> <p>This has been delivered throughout; see deliverables section.</p>	
<b>Section Three: Institutional &amp; Project Partner Issues</b>	

Each partner has filled in a separate report, appended, under four common headings to cover institutional-specific issues and progress. Due to ongoing staffing problems, LSE has not filled in a partner report at this time. See risks section.

The Consortium Agreement was finalised at Edinburgh and initially circulated to the partners. A few minor changes are required, but we expect to get this signed by all partners shortly.

## Section Four: Outputs and Deliverables

### Publications and Presentations <http://www.disc-uk.org/publications.html>

- RICE, R. and GIBBS, H. (2008) [From Open Access to Open Data](#). Third International Conference on Open Repositories 2008, 1-4 April 2008, Southampton, UK.
- MACDONALD, S. and MARTINEZ, L. (2008) Libraries in the Converging Worlds of Open Data, E-research, and Web 2.0. *Online*, 32 (2), 36-40.
- MARTINEZ, L. (2008) [The Data Documentation Initiative \(DDI\) and Institutional Repositories](#). DISC-UK.
- MACDONALD, S. (2008) [DISC-UK DataShare: Web 2.0 Data Visualisation Tools: Part 1 - Numeric Data](#). DISC-UK.
- RICE, R., GREEN, A., and RUMSEY, S. (2007) [DISC-UK DataShare: Building capacity for institutional data repositories](#) (presented at the [3rd International Digital Curation Centre conference](#), Washington, DC, December 11-13 2008).

Additionally, a number of new key references, e.g. from the *State-of-the-Art Review*, have been added under [References on Data Sharing](#) on the above web page.

A number of presentations were given to the project team at our Face2Face meeting in February, 2008 (see activities section). Ann Green has kindly agreed for the following two presentations to be shared publicly, and are linked from our Deliverables web page. [Data Documentation Initiative \(DDI\)](#), by Ann Green [PDF]. [Guidelines and Tools for Repository Planning and Assessment](#), by Ann Green [PDF]. (Other presentations and notes posted on team filestore.)

### Collective Intelligence (Web 2.0 information sharing) <http://www.disc-uk.org/collective.html>

- *April, 2008*: Forty-six annotated bookmarks on project themes added to Faves account during reporting period (106 total). These can be retrieved as a dynamic bibliography using a **live tag cloud** on the collective intelligence page above.
- *April, 2008*: Report back from RSP Repository Services Day, Nottingham, by Theo Andrew [blog entry].
- *March, 2008*: [A UK Research Data Management Forum](#), report from Manchester, by Luis Martinez [blog entry].
- *April, 2008*: [Report back from Open Knowledge conference](#), LSE, 15 March 2008 by Robin Rice [blog entry].
- *March, 2008*: [Blogs for Statistics and Data in the Social Sciences and beyond](#) - As part of the ESRC Festival of Social Science 7th - 16th March (with an emphasis on open data), by Stuart Macdonald.
- *March, 2008*: [What's Different About Data?](#), by Robin Rice [blog entry].
- *February 2008*: Selected **RSS newsfeeds** added under related sites.

### **Section Five: Outcomes and Lessons Learned**

So far, these remain about the same as at the last reporting period.

### **Section Six: Evaluation**

Following requests to possible external evaluators by the project manager, a proposal was received by Sheila Anderson, Kings College London, for external formative and summative evaluation and accepted pending revisions. Once we receive the revised proposal, we'll inform JISC of the evaluation plan..

### **Section Seven: Dissemination**

*April, 2008:* The project manager attended the Significant Properties Workshop by the DCC at the BL. Contacts were made with Richard Davis, ULCC, and the Cambridge Repository Manager, Elin Stangeland.

*April, 2008:* A number of project partners attended Open Repositories 2008 in Southampton. Ben O'Steen (Oxford) was on the winning team of the Repository Challenge. Luis Martinez and Stuart Macdonald have submitted a report of the event to CILIPS for publication. A number of useful contacts from the UK and other countries were made. A poster session was given and leaflets were disseminated.

*April, 2008:* Edinburgh hosted Australian and New Zealand repository managers on study visits to the UK: Maude Frances, UNSW and Leonie Hayes, University of Auckland.

*March, 2008:* Both Robin Rice and Luis Martinez attended the UK Research Data Management Forum inaugural workshop, organised by the DCC and RIN.

*March, 2008:* A 'lightning presentation' was given at the Open Knowledge Conference, LSE, 15 March 2008 on DataShare by Robin Rice [webcast at <http://www.archive.org/details/OkCon2008> (final MP3 file, about half-way through)]. Leaflets were also disseminated here and at the Open Repositories conference. She discussed the project and related issues with Dave Puplett and Jenny Brace (LSE); Jordan Hatcher, co-author of the Open Data License; and Jonathan Gray, Open Knowledge Foundation.

*March, 2008:* Jane Roberts (Oxford) attended the ESDS event on Managing and sharing research data in Leeds. A blog entry reporting on the event will be written.

*February, 2008:* Wendy White (Southampton) went on a study visit to Japan as part of another JISC Repositories project (KULTUR).

*January, 2008:* Tanvi Desai (LSE) attended the ESDS Government meeting on the new Statistics Act in London. A briefing paper on the implications of the Act will be written.

*February, 2008:* The Co-Project Director met with Michael Jubb from RIN and discussed the project with him while he was visiting Edinburgh University.

*December, 2007:* At the DCC conference in Washington, DC, the project manager discussed the project with others doing related work in the US: Jake Carlson, Purdue University; Gail Steinhart, Cornell University Library, Mann Library Data Staging Repository (DataStaR); Katherine McNeill-Harman, MIT Libraries, DSpace and the Harvard-MIT Data Center; Tim DiLauro, Digital Library Architect, Johns Hopkins University; Gretchen Gano, Assistant Curator, Librarian for Public Administration & Government Information, New York University Libraries.

*December, 2007:* Stuart Macdonald (Edinburgh) gave a [poster session](#) at the ESDS International Conference, 3 December on *EDINA Repository Projects*.

*December, 2007:* An EDINA Newline article, "DISC-UK DataShare project to demystify institutional data," plus a later EDINA news item announcing deliverables brought attention from Peter Suber's Open Access News and other quarters. DISC-UK authors of reports fielded some queries from as far as the US, New Zealand and Australia.

*November, 2008:* The project manager attended the DCC workshop in Glasgow: Legal environment of data curation and discussed the Open Data License with Jordan Hatcher.

*November, 2008:* The project manager attended the JISC Programme Meeting, along with some other project partners.

## Section Eight: Risks, Issues and Challenges

### LSE and DataShare

London School of Economics and Political Science has lost one key member of Library staff, Luis Martinez, the Data Librarian at LSE Library, during the first year of the DataShare Project. Luis Martinez left LSE in December 2007 to take up a post at the University of Oxford. Despite two rounds of recruitment in early 2008, LSE has not been able to fill the position of Data Library Manager, as of April 2008.

LSE has also experienced difficulties in recruiting to the Technical Officer Project post. There have been two formal rounds of recruitment for this post as well as a circulation to suitable agencies. However it has not as of April 2008 been possible to recruit.

LSE is currently reviewing these two posts and considering how best to fulfil its commitments to the Project, in consultation with the Project Manager.

## Section Nine: Collaboration and Support

We have had more contacts from people outside of the SUE programme (and indeed, outside of the UK) although we share some common interests (particularly software-related) with other repository projects. We are not sure if this is because the SUE projects are busy with their current deliverables, or if this reflects on the current interest levels in using repository services for research data in the UK.

We (DISC-UK) would be interested in hosting a data-related event or programme meeting, to coincide with the 25<sup>th</sup> anniversary of the Edinburgh University Data Library this year, though we do not know how much attendance we can get from SUE projects due to the observation above.

*March, 2008:* The project manager was invited to join a list and attend meetings organised by Neil Jacobs at JISC to coordinate common activities, such as training events. The group consists of:

- Alma Swan (data curation skills / careers study)
- Joy Davidson (DCC data curation summer schools)
- Seamus Ross (Data Audit Framework)
- Robin Rice (DISC-UK Datashare project)
- Jean Sykes, John Milner, Martin Sewell (UK Research Data Service (UKRDS) feasibility study)
- Neil Beagrie (Data preservation costs study)
- Graham Pryor (DCC)
- Martin Lewis (CURL / project steering groups)
- Matthew Dovey, Neil Grindley, Alison Turner (JISC)

Neil has posted a [blog item](#) about this group.

The partners are participating in other projects and we are either actively collaborating or looking for synergies with [VIF](#) (LSE), [FAR](#) (LSE), [BID](#) (Oxford), [Scoping Digital Repository Services for Research Data Management](#) (Oxford), ShareGeo (Edinburgh), [Data Audit Framework Implementation](#) (Edinburgh), and [Kultur](#) (Southampton).

## Section Eleven: Next Steps

*In this section you should very briefly list the activities planned and/ other information of relevance for the next stage of the project.*

- Ann Green will be compiling documents from each site and revising a questionnaire to determine readiness for repositories to accept submissions of research data, based on the sources covered in her presentation to the face to face meeting (see Outputs section); this will be used by the partner sites over the next few months to work on Workpackage 2 (enhancing repositories) and disseminated for external use.
- Following successful completion of WP 2 deliverables – getting IRs ready for ingest of datasets, the advanced WP 3 deliverables will be pursued.
- LSE will be meeting with the project manager to determine its plan for the rest of the project, and whether funding and deliverables need to be rearranged. The programme manager will be kept informed of developments and consulted if need be.
- The project team will determine if and when a programme meeting type of event will be held in Edinburgh during 2008.
- Upcoming deliverables are in process that will be posted on the website as they are published.
- Case studies are being written to document particular interactions with researchers and unique problems, such as migrating a Microsoft Outlook database to a generic format.
- A poster session was accepted for the National Centre for e-Social Science in July, by Stuart Macdonald
- A poster session was submitted to the Dublin Core conference in September, by Robin Rice.
- A presentation will be given about DataShare at the IASSIST conference in May in Stanford by Robin Rice, and about Data Mashups by Stuart Macdonald, and about the Development of Remote Access Systems by Tanvi Desai.
- A paper will be submitted to the annual DCC conference, December, 2008.

### Checklist:

Before you return this report:

- Ensure that your project webpage on the JISC site is up to date and contains the correct information. Attach details of any required amendments to this report. Project webpages can be found from: <http://www.jisc.ac.uk/whatwedo/projects.aspx>
- If there have been any changes to the original project plan and/or work packages, ensure that amended copies of the relevant sections of your project plan are attached to this report.

## DataShare: Edinburgh Progress Report, April 2008

### Advocacy and Promotion

#### Joining up Repositories within IS (Information Services)

Edinburgh DataShare - the name of the pilot repository set up by the Data Library as part of the JISC project - is operated within a different IS division and runs on a different server than the pre-existing eprints repository run by the Library (ERA). Both use DSpace and staff have met to discuss how to harmonise the two. Presently ERA (Edinburgh Research Archive) and the Publications Repository (which was set up to administer the RAE submission for the University) constitute two different repositories and two different services as well. A development plan is proposed to bring the two repositories together to form a single service. In this service environment the Publications Repository would aim to be a complete record of research outputs for the University and provide management of the content with access limited to the University of Edinburgh. ERA will include all content that is open access and be the public view of research at the University.

#### Outreach and User Engagement

- As part of the Publications Repository advocacy and outreach activities Edinburgh DataShare has been provided with the opportunity to contribute to such events, the first of which was a presentation to the School of Education Administrators on 20/3/08 by Stuart Macdonald. This followed a discussion that Robin Rice had with Dr Linda Croxford from the School of Education regarding the potential acquisition of research datasets for ingestion into Edinburgh DataShare. Edinburgh DataShare will also be presented to School and Research Administrators for the Humanities and Social Science College at a larger event planned for 24/4/08. Subsequent events will follow for the Medicine and Veterinary Medicine, and the Science and Engineering Colleges in due course.
- Peter Burnhill introduced Edinburgh DataShare (as part of a larger presentation) to University of Edinburgh IT Professionals Forum (22/4/08).
- Research data users from the School of Scottish Studies and Scottish Centre for Crime and Justice Research have presented to us data for inclusion in DataShare. This has engendered much discussion about issues of copyright, ownership, formatting, preservation etc. After meeting with Steve Boardman from the School of Scottish Studies indicated that there could be other similar research databases requiring a home. This will be followed up.
- Edinburgh DataShare now has links from the Information Services website. This has resulted in a small number of enquiries pertaining to the deposition of research datasets.
- Edinburgh DataShare was introduced to postgraduate and academic staff at EUCS Training workshops (Using EDINA Digimap and Visualising Data using ArcGIS).
- Edinburgh DataShare has been showcased to its academic community in the Bulletin of IT Services (BITs) in addition to the EDINA Newline quarterly (December 2007).
- The Edinburgh University Quantitative Social Science Network (EQuSS-N) has been established to link together people from across a range of Schools and disciplines who are interested or engaged in research or teaching that involves quantitative methods. This network has been introduced to the concepts behind DataShare from an institutional perspective.

#### Synergies

IS recently secured funding from the JISC to implement a Data Audit Framework. This will be conducted across targeted departments the results of which will be fed back into the DataShare project in terms of the identification of data users whose materials may be suitable for ingestion into the DataShare repository. A local steering committee for the DAF will help to ensure that the work will be relevant to IS strategic aims, which in turn will help the exit strategy for the local DataShare project. Robin Rice is project director with Cuna Ekmeckioglu from IS acting as project manager. Cuna has been added to the DataShare JISCmail list and will in effect be part of the DataShare team from now on.

ShareGeo is a newly funded short-term JISC project (May-September) which will migrate the GRADE DSpace repository set up by EDINA as part of the Digital Repositories Programme into a new version of DSpace and to operate within the secure environment of Digimap (due to the problems identified with sharing data derived from OS map data in GRADE). Today the repository has over 160 deposited datasets and over 150 registered users.

Mary Robinson from SHERPA gave a talk on the DRIVER Guidelines to EDINA staff including the Depot, JORUM, and the DataShare team, 25/1/08.

Interested parties from Information Services were invited to meet with two Repository Managers on study visits from their respective universities, Leonie Hayes (University of Auckland - 28/3/08) and Maude Francis (University of New South Wales - 11/4/08) who made contact with the DataShare project manager. This offered the opportunity to learn about related developments in Australia, and importantly, for repository staff across IS to come together to discuss issues in common and to get to know each other over a meal.

## **Policy and Strategy**

Some strategic directions will be informed by our inter-working with other related services and projects, as described in the Advocacy & Promotion section.

Unlike the other partners, Edinburgh DataShare is more of a start-up repository, because it is separate from ERA. Therefore, we have a short time to implement everything we'd like to have operational by the end of the project. The decision to call this a pilot repository was taken because it is unlikely we will have everything running perfectly by the end of the project. We would like to launch a more stable service around the time of the end of the JISC project, but this will depend on our ability to embed funding for the service within existing services in IS, such as the Data Library. It will also depend on the perceived value of the service by local stakeholders.

Many policy decisions were identified at the original kick-off meetings, such as scope, size, metadata schemas, preservation, access & licensing, quality assurance, self-deposit vs. assisted deposit. We explored many of these questions more deeply at the face to face meeting in February. For Edinburgh DataShare, we may have to decide some of these issues as we build the repository and negotiate with depositors, taking an informed but organic approach. By the time of the next report we hope that we will have resolved most of these policy questions and have them well-documented within the service.

## **Technical and Metadata Enhancements to IRs**

### **Installation and Configuration of DSpace**

DSpace Version 1.4.2 was installed on a Solaris server using Apache Tomcat web server and a PostgreSQL database.

### **Look and Feel**

The repository look and feel was modified to match the Edinburgh Research Archive's appearance.

### **User Authentication**

DSpace was modified to use the University of Edinburgh's EASE authentication system. EASE is based on the Cosign Single Sign-On system which is not supported by the DSpace authentication stack. In consequence, some core modifications were made to DSpace to allow EASE to be used as the DataShare authentication system. In addition, an LDAP server is used to simplify the registration process - filling out some registration fields.

### **Statistics Enhancement**

The default statistic package was configured but deemed not to tell the repository managers what bitstreams were being downloaded and who was downloading them. The statistical package was modified to display the number of times an individual bitstream was downloaded (either monthly or in-total) and the user could select a hyperlink to view a table containing the ip and email address (if logged in) of the downloader, and time of the download.

### **Controlled Vocabulary**

To support standard metadata values during item deposits, the Joint Academic Coding System (JACS) was chosen as a controlled vocabulary standard. As the standard is distributed as a comma separated file some processing of the file was required to generate the XML format expected by DSpace.

### **Handle Server**

To support long-term URI identification and resolution, the Handle.net handle server was installed and configured in DSpace.



## **Data Harvesting**

DSpace was configured to allow OAI-PMH harvesting and the service was registered with OpenDOAR.

## **Upgrade to DSpace 1.5**

The re-organisation of the DSpace codebase, build system and web development framework in version 1.5 forced a re-evaluation of some of what was implemented above.

- The decision to port DataShare DSpace to the new XML User Interface (Manakin) made the JSP statistics enhancements obsolete, as Manakin 1.5 has no statistical package.
- The Edinburgh DSpace look and feel was re-written as a Manakin Theme
- The EASE/cosign authentication code had to be modified (as authentication was no longer initiated by a Servlet Filter).

The live pilot instance is currently running DSpace 1.5

## **Dublin Core**

A decision was made to use qualified Dublin Core in a way that enhanced the information about the dataset as much as possible; e.g. use of GeoNames to describe geographic coverage, standard ways of inputting date, and use of terms such as Creator to be compatible with the DDI (Data Documentation Initiative). Each metadata field chosen is being reviewed for quality control and then will be implemented in the interface.

## **Workflow**

The decisions about metadata fields may affect the workflow interface, for example the order certain fields appear, and whether certain questions need to be asked for a field to appear (such as whether there's a published version elsewhere). This still needs to be done, within the confines of the new DSpace implementation, and any help or hint text still needs to be written to prompt the user how to respond. Decisions about required fields have yet to be taken as well.

## **Assessing Impact**

We feel it is a bit premature to assess impact within our user community because we are still in set-up mode, and there are no statistics to count yet. On the other hand, communication with researchers has been very valuable in understanding their needs. There is much going on within IS that we would like to be cooperating with, that impacts on users' needs for assistance with data management. By carrying out the JISC Data Audit Framework implementation within the University, we are building on a research computing survey that began to differentiate types of data users and their computing needs.

Again, there may be more that was said under Advocacy which provides anecdotal evidence of our impact so far.

## DataShare: Oxford Progress Report - April 2008

### Advocacy and Promotion

A former project partner at the LSE, Luis Martinez, is now in post at the Oxford e-Research Centre as Digital Repositories Research Coordinator, and is working alongside the Oxford DataShare team. His project - *Scoping Digital Repository Services for Research Data Management* - is closely related to the purposes of DataShare. The project is now at the stage of conducting interviews with researchers across disciplines in which he will learn how they manage their data and identify gaps in the provision of services. From her knowledge of the activities and practices of social science researchers and data collectors, Jane Roberts has provided suitable and willing candidates for interview. A report with recommendations will be produced in July.

Luis presented a seminar *The management of research data in digital repositories* where he introduced the concept of data management and highlighted some benefits and challenges of managing research data. He discussed repositories and how they can be used for storing and disseminating data, showing some examples of projects already happening at Oxford (including DataShare) and internationally.

Sally Rumsey and software developer Ben O'Steen presented two seminars aimed at all those dealing with research materials (including authors, administrative and IT staff) - both were well attended. There has been a senior level meeting about the use of ORA for the Research Excellence Framework. There have been visits and presentations in a number of academic departments and research institutes to raise awareness of ORA; several now intend to use ORA for their research materials although we are yet to see the results of these decisions.

Luis and Sally are taking part in the UK RDS (Research Data Service) consultations, for which Oxford is a case study.

Luis will be participating in the Digital Curation Centre's process of developing a week-long course on digital curation. Aimed at bench scientists, computing scientists, information specialists, and data providers the course will focus on the practical, rather than the theoretical, and will contribute towards bridging the gap that currently exists between a general awareness of curation and preservation issues amongst the information services and scientific community, and their ability to practically address digital curation and preservation challenges within their actual working environment.

### Policy and Strategy

ORA policies are available at [http://www.ouls.ox.ac.uk/ora/ora\\_documents2/ora\\_policies](http://www.ouls.ox.ac.uk/ora/ora_documents2/ora_policies). The scope of ORA for use for research data has not yet been set, and therefore the specific policies for handling data are yet to be tackled.

### Technical and Metadata Enhancements to IRs

#### Repository development

ORA has a new interface (release date April 2008). In addition to the fundamental functions of search, browse and access it includes a number of new features:

- faceted search and browse
- visible RDF relationships (non-hierarchical which can be used for relating digital objects within and without ORA e.g. article to dataset)
- example relationships for books (i.e. chapter relating to book/work and image (page) relating to chapter which relates to book/work)
- conferences with related papers/materials
- OAI-ORE resource maps

We have written a draft risk analysis using a section of the DRAMBURA toolkit. This will be further developed for the ORA end of Development Phase report (to be completed summer 2008 and presented to the ORA Steering Group in autumn 2008).

Oxford social science data library already has a small collection of locally produced datasets available for download from its website. These will be used as test cases for ingest into ORA.

### **Digital preservation**

ORA will shortly be beginning trials with the SHERPA DP2 project investigating the preservation of repository content within a centralised dedicated preservation repository. Although our content will be harvested for the purposes of the trial, in practice we would expect to be a preservation service provider to others.

At OR08, the PRESERV2 development team demonstrated the wholesale transfer of repository content between FEDORA (ORA) and EPRINTS (OR08 conference repository) repositories using the OAI-ORE protocol and winning the JISC CRIG Code Challenge in the process.

### **Bridging the Interoperability Divide (BID) project**

Ben O'Steen has made the ORA code for harvesting material and metadata from other repositories available in Google Code. Having developed a series of code libraries which make repository operations accessible to a Python environment, the actual harvesting script is quite simple and compact. Further investigation of the Numerical Metadata standard and DDI leads us to believe that we will have to support multiple dataset metadata schemas in the near term as there are insufficient standards in place at the current time. These will be added to the Google Code posting in due course.

### **Assessing Impact**

The advocacy work for the Oxford repository carried out to date has shown much enthusiasm from potential users from many disciplines. We are not yet in a position to assess the actual impact of the proposed enhancements for data ingest on our research community.

## DataShare: Southampton Progress Report April 2008

### Advocacy & Promotion

Advocacy and promotion has continued to be the focus of activity at Southampton. Whilst continuing discussions with staff in the School of Social Sciences, we also took the decision to broaden our scope to other subject areas. These include maths, social work, health, education, history, music, film studies and archaeology.

Promotion of DataShare has taken place with additional assistance from Subject Librarians working with researchers in the fields listed above. Approaches have varied according to the nature of respective Schools/Departments but include emails to staff lists, discussions with key individuals and discussing DataShare at meetings.

These efforts have resulted in the establishment of contact with a retiring Archaeologist who is keen to find storage for and disseminate both her publications and data. Her data is unsuitable for AHDS and this, therefore, presents precisely the type of situation that DataShare set out to address.

Contact has also been made with a Professor of Film Studies who is developing a database of German-Speaking Emigres in British Cinema. The database is to be made freely available from a web interface and is currently work-in-progress.

In other areas researchers feel that they do not hold suitable data that has not been deposited elsewhere. Archaeology and history, for example, continue to have specialised data centres despite the withdrawal of funding from AHDS. Having said this, Historians at Southampton do not appear to have been engaged with AHDS because they do not appear to generate suitable data. In social work a new research centre has recently been established. The Centre Head was receptive to DataShare but felt she could not be more involved at this very early stage of the Centre's life. In Music researchers are more concerned with video and audio archiving.

Returning to the School of Social Sciences, the Project Officer has been invited to attend the next School Research Committee meeting which is attended by all Departmental Heads and Heads of research centres. The Deputy Head of School for Research has also agreed to set up a meeting with final year PhD students to discuss DataShare in relation to their research data. A meeting will also be set up with the ESRC funded National Centre for Research Methods which is based at Southampton and also the work being done by the related Restore project. <http://www.ncrm.ac.uk/research/other/restore/> Wendy White is on the Restore advisory group.

### Strategy and Policy

Issues raised by DataShare have contributed towards the establishment of a working group charged with the development of an institutional preservation policy. Chaired by Mark Brown, the group reports to the University Research Committee. The outcome of this work almost certainly impact on the preservation of data at Southampton.

The database of German-Speaking Emigres in British Cinema, mentioned above, raises a number of policy and practical questions.

- Should we take works-in-progress and, if so, how would these be managed?
- Do we wish to duplicate work that is freely available elsewhere?
- Should we take the underlying database or consider website archiving?

## **Technical & Metadata Enhancements to IRs**

e-Prints Soton currently includes an embargo facility, allowing researchers to deposit data at a convenient time whilst controlling its release. An upgrade to e-Prints 3.1 is scheduled for December. Amongst other benefits, this will introduce improved time-stamping and track edit facilities.

Metadata developments are being carried out in collaboration with the DataShare partners. Discussions held at the face-to-face meetings in February were informative and work currently being carried out by Ann Green should assist in the decision making process. One issue we expect to address is the balance of metadata between file and item levels.

The e-Prints Soton software developer has scheduled time to work on DataShare developments before October. We hope to have examples of data from researchers in social sciences and archaeology to use in the technical and metadata developments, even if it is not possible to make these datasets open access.

## **Assessing Impact**

Impact will be assessed through the external consultants working across the project institutions.