



Progress Report

Project Name	DISC-UK DataShare (EDINA)
Project Website	http://www.disc-uk.org/datashare.html
Report compiled by	Robin Rice
Reporting period	Apr 2008 – Oct 2008

Section One: Summary

The project has changed some of its deliverables following the change in LSE's status to an associate partner due to staffing shortages; this includes a greater emphasis on data audits at each of the remaining partners to reach out to users earlier in the data lifecycle and to better meet their needs for support in data management. Edinburgh participated in the Data Audit Framework Development project led by HATII/DCC at University of Glasgow and conducted its own DAF Implementation project, both funded by JISC, and led by Robin Rice, with a new team member, Cuna Ekmekcioglu.

The project members have continued to engage with contacts at UK and international institutions – especially in the US and Australia - who are building services for data sharing. The project team has participated in professional development activities, and disseminated deliverables at conferences, in articles, and through their website and blog. A briefing paper on geo-spatial Web 2.0 visualisation tools was written. The findings from Oxford's *Scoping digital repository services for research data management* project were disseminated. Several project team members participated in the Edinburgh Repository Fringe, along with peer projects amongst the partners, e.g. Kultur and EdShare (Southampton), ShareGeo and the Depot (EDINA), and the CRIG International Roadshow (Oxford). An article in *Online* by Luis Martinez Uribe and Stuart Macdonald and an interview in *CILIPS Update* brought attention to the profession of data librarians, which was further amplified by the recent JISC-commissioned report by Key Perspectives, *The Skills, Role and Career Structure of Data Scientists*.

The partners have created and received peer review on a Dublin Core based metadata schema for datasets in DSpace and EPrints, worked on procedures for storing and preserving databases, and have developed a content model for a database of sound files in Fedora. The Edinburgh DataShare repository was soft-launched, with an option for depositors to append the open data license developed by the Open Data Commons.

Section Two: Activities and Progress

WP 1: Project Management

Regular communication was maintained through the project list and through telecon meetings, with a summertime hiatus. Opportunities for professional development were raised to project staff, particularly ones from the SUE strand; those who attended events were requested to write blog entries. The website was maintained as was communication with JISC throughout. The project manager attended a JISC meeting called by Neil Jacobs to bring together related data projects in July and met with Neil at an EDINA visit in May. The external evaluator for the project was selected (Sheila Anderson, Kings College London) and the proposal agreed. Negotiations with partners took place regarding LSE's change to associate partner and the remaining funds and deliverables, resulting in proposals from Southampton and Oxford to take up data audit work as piloted by Edinburgh through a DAF implementation project. Edinburgh to extend its data audit work beyond the life of the DAFI project by continuing under DataShare. The original workpackages document has been updated to reflect these changes and is available on the project website's deliverables page.

WP2: Institutional Repository Development

Please see individual progress reports.

WP3: Technology

A number of developments on the technology front have been made at the partner sites.

Edinburgh made a number of enhancements to DSpace, including an embargo feature, automatic account creation based on the single-sign-on system used for staff and students (EASE), automatic system tests were written, and a spam prevention measure was put in place for forms (see Edinburgh report for details). A procedure was developed for saving Access databases in plain text tables with an SQL schema for importing into any database by end-users. A metadata schema was put in place which was disseminated on the website, blog, and at the Dublin Core 2008 conference as a poster - http://dc2008.de/wp-content/uploads/2008/10/12_rice_poster.pdf - as a potential for a generic application profile for datasets. An option for depositors to append an Open Data License for end-users was installed.

Oxford has implemented a content model for deposit of their first large database. See Oxford progress report; also Ben O'Steen's blog post, Less Talk, More Code: Modelling and storing a phonetics database inside a store - <http://oxfordrepo.blogspot.com/2008/10/modelling-and-storing-phonetics.html>.

Southampton has specified the metadata fields for a new dataset *type* to be incorporated during a future upgrade to EPrints 3.1, which has synchronicity with the Edinburgh schema but also pre-existing fields in the Soton repository.

WP4: Outreach and Dissemination

This has been delivered throughout; see Outputs and deliverables, section 4.

Section Three: Institutional & Project Partner Issues

Each partner has filled in a separate report, appended, under four common headings to cover institutional-specific issues and progress.

The Consortium Agreement needed to be revised and a new budget appended following LSE's departure as a full partner. This is expected to be signed and returned to JISC at the same time as this report.

Section Four: Outputs and Deliverables

October 2008: Revised [workpackages](#) reflecting London School of Economics change to Associate Partner status and the addition of Data Audit Framework tasks at the three partner institutions.

October 2008: **Edinburgh DataShare** repository announced to be [open for business](#) in the University IT newsletter, BITS.

October 2008: London School of Economics, as associate partner in DISC-UK DataShare, [announces new data librarian in post](#) [blog entry].

October 2008: Stuart Macdonald and Luis Martinez Uribe each attended the week-long DCC 101 training course at the National e-Science Centre in Edinburgh.

September 2008: Robin Rice presented a poster at the International Conference on Dublin Core and Metadata Applications (DC-2008) in Berlin, on [Applying DC to Institutional Data Repositories](#), [PDF]. Additionally she gave an invited brief presentation at the [Metadata for Scientific Datasets Workshop](#) held during the conference.

September 2008: Briefing paper. [Data Visualisation Tools: Part 2 - Spatial Data in a Web 2.0 Environment and Beyond](#), by Stuart Macdonald [PDF]. Minor updates October 2008.

September, 2008: [New Report - The Skills, Role and Career Structure of Data Scientists](#), by Luis Martinez [blog entry].

September, 2008: [International Census Microdata Conference: Some resources recommended by speakers and JISC - Rights and Repositories Programme Meeting](#), Tanvi Desai [2 blog entries].

August, 2008: [Notes from the Edinburgh Repository Fringe](#), by Stuart Macdonald [3 blog entries].

August, 2008: Stuart Macdonald presented [Open Data, Open Minds - Web 2.0 data visualisation tools at Edinburgh Repository Fringe 2008](#), Playfair Library, University of Edinburgh, 1 August 2008.

July, 2008: Luis Martinez Uribe presented the findings of the Oxford Research Data Management Project - [Scoping digital repository services for research data management](#) at [Edinburgh Repository Fringe 2008](#), Playfair Library, University of Edinburgh, 31 July 2008.

July, 2008: Robin Rice gave a DISC-UK DataShare project 'Soap-box' presentation - Opening the Doors: Minding the Gaps at [Edinburgh Repository Fringe 2008](#), Playfair Library, University of Edinburgh, 31 July 2008.

July, 2008: Stuart Macdonald presented [DISC-UK DataShare](#) at the [MIMAS Open Forum](#), University of Manchester, 9 July 2008.

July, 2008: Robin Rice contributed an article about the project to a special issue on digital repositories in *ALISS Quarterly* (3)4.

July, 2008: [DSpace Metadata Schema for Edinburgh DataShare](#) - Ver. 1, 3 July 2008 by Robin Rice, Stuart Macdonald and George Hamilton [PDF].

July, 2008: [DSpace for Data](#), 3 July 2008 by Robin Rice [blog entry].

June, 2008: [DISC-UK goes to Silicon Valley for IASSIST 08](#), 27 June 2008 by Jane Roberts [blog entry].

June, 2008: Robin Rice presented [DataShare and Data Audit Framework projects at Edinburgh](#) at the [DCC/Edinburgh e-Science Collaborative Workshop](#), National e-Science Centre, University of Edinburgh, 12 June 2008 and at the [Research Data Management Workshop](#), University of Oxford, 13 June, 2008.

June, 2008: [Report from the RELU/UKDA Anonymity and Consent workshop](#), Edinburgh University, 17 June 2008 by Stuart Macdonald [blog entry].

June, 2008: Interview with Luis Martinez-Urbe and Stuart Macdonald published in *CILIP Update*; June 2008. [Data Librarianship - a gap in the market](#).

June, 2008: Open Repositories 2008 Conference Report published in *CILIP Update*; June 2008, by Stuart Macdonald and Luis Martinez-Urbe.

May, 2008: Tanvi Desai (LSE), Robin Rice (Edinburgh), Stuart Macdonald (Edinburgh) gave presentations at the [International Association for Social Science Information Service and Technology \(IASSIST\) annual conference](#), Stanford University, US, 28-30 May 2008.

Section Five: Outcomes and Lessons Learned

We are in the midst of a time that is seeing a lot of attention on both institutional repository development and on increased pressure for research data to be shared, curated, preserved and re-used. Our exemplars do seem to be helping to build confidence that this is something that institutions can do; on the other hand, it is not easy to get deposits at a time when the open access argument for research papers has not yet taken a strong hold within the research disciplines. By embracing the Data Audit Framework activity, we are able to gain access to researchers earlier in their research process (cycle), and also to offer something that is seen as beneficial to them; help to improve practice at normal data management, with or without sharing. Gaining leverage from this will take some time and the critical mass will not reach our repositories until after the project is over. At our next face to face meeting we will spend time talking about our sustainability plans and whether the experiment with data repositories will continue at each institution.

Section Six: Evaluation

A proposal was accepted for Sheila Anderson, Kings College London, to be the external evaluator. The evaluation plan was submitted to the programme manager and is appended to this report.

Section Seven: Dissemination

A full list of activities is given above. Staff have blogged about events they have attended for professional development; reports and other material useful to the project has been bookmarked and annotated for a living bibliography (tag cloud) on our [collective intelligence](#) web page. There are 134 references there now. Staff have also given presentations and posters at relevant events; and one project briefing paper has been written in the period: [Data Visualisation Tools: Part 2 - Spatial Data in a Web 2.0 Environment and Beyond](#), by Stuart Macdonald. The project manager engaged with other JISC projects at the JISC Innovation Forum at Keele in July, the Oxford Data Curation meeting in June and several project team members participated in the Edinburgh Repository Fringe. Visitors included Herbert von de Soempel (Los Alamos Labs), Paula Callan (Queensland University of Technology), Rowan Brownlee (University of Sidney), Brian Flaherty (University of Auckland), and Andrew Treloar and Andrew Groenewegen (Monash University/ANDS).

Section Eight: Risks, Issues and Challenges

The project has withstood the withdrawal of LSE as a full partner; the stool with three legs remains standing.

A number of alternative solutions for data sharing have arisen; UKRDS, STORE, for example. It's not clear how institutional data repositories "fit in" to the landscape.

None of the partners have acquired a critical mass of datasets yet, so we may not have usage statistics as an output. Each partner needs to consider whether this is due to slow starts and uphill learning curves, or something more pernicious.

Over the next period, each partner will need to determine its sustainability plan or exit strategy.

Section Nine: Collaboration and Support

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

We have collaborated with the DAF Development project (DCC/HATII). We have begun discussions with the RSP about an upcoming deliverable on policy development for datasets. We have been in contact with a number of like-minded professionals in the UK and around the world through various conferences. The project manager has been asked to help form a dataset interest group in the DCMI following the Dublin Core 2008 conference. Staff at UKOLN and York gave us advice on our DC metadata schema. We remain in contact with past visitors (see section 7) and are open to collaboration opportunities with them.

Budget Report - omitted

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.

- The Consortium Agreement with a revised budget and workpackages has been circulated for signature and will be submitted to JISC at the end of October.
- The project manager will give an invited presentation on data librarians at the RIN/DCC Data Management Forum in Manchester in November.
- The project manager will co-present at the Innovation Fair at the SPARC Digital Repositories meeting in Baltimore in November on “Geospatially enabling DSpace repositories: standards-compliant ways to georeference items in DSpace.” The project manager of ShareGeo is the co-presenter.
- Luis Martinez and Stuart Macdonald will each give a presentation at the Madroño Data Repositories Seminar 2008 in Madrid, Spain in November, as invited speakers.
- Robin Rice has submitted an article to the *IQ (IASSIST Quarterly)* which will be published in the autumn, 2008.
- Ann Green is leading on a deliverable to develop policies for repositories incorporating datasets, based on a distillation of state of the art standards and documents, and partners’ existing policies. This may lead into a training session proposal at the next IASSIST conference, depending on the outcome of the face to face meeting in December.
- The project manager is seeking a legal expert to draft creative commons-like licenses for data repositories. (Edinburgh DataShare currently offers the Public Domain Dedication and License but would like to offer other options as well. Creative Commons licenses are only suited to copyrightable material, not data.)
- The formative evaluation will take place over the next month or two and will be reported at the next face to face meeting. The project manager will also participate in the programme-wide evaluation through a telephone interview.
- A face to face meeting is planned for 4 December in Edinburgh. This is preceded by the DCC conference in Edinburgh and will be followed by the Edinburgh University Data Library’s 25th anniversary celebration on 5 December, with a *Symposium on Institutional Data Services*.
- Study visits are being planned by the project manager for institutions in Australia that have developed data curation and support services. These will be written up and blogged.
- The final evaluation report will be submitted following interviews as described in the evaluation plan.

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.

Appendices:

Evaluation plan

Partner progress reports

Evaluation Proposal for the DISC-UK DataShare Project

Sheila Anderson, Centre for e-Research, King's College London, September 2008

Introduction

The DataShare Project has set aside a sum of money to undertake an evaluation of the project. In an initial discussion with Robin Rice we identified three key elements that the evaluation would ideally cover:

- The impact of the project on the individual project partners
- The wider context as a collaborative consortium seeking to share and learn from each other
- The relevance of the project deliverables and outcomes to the wider HE community

The preference is for an evaluation that would be both formative and summative.

This document therefore proposes a two-stage evaluation process:

- A 'light touch' formative evaluation with the project partners to gather evidence on their expectations and the impact of the project activities to date
- An in-depth summative evaluation that will involve interviews with the project partners and a small group of stakeholders who will be asked to comment on the relevance of the project deliverables to the wider HE community

Methodology

The methodology we would propose to use is as follows:

- Telephone interviews with each of the three partners which will concentrate on formative issues and experiences. These will take place in late October and November and a short report summarising the interviews will be presented to the partners in December.
- In-depth interviews with each of the three partners on the summative issues, experiences and final project outcomes. These interviews will take place in March 09 at the end of the project in order to gain maximum value from the interviews. The evidence gathered will be used to create case studies reflecting the experience and impact for each partner.
- Analysis of the evidence gathered during the summative and formative interviews to identify and collate information on the challenges and successes of the collaborative experience and the value-added offered (or not of course).
- Interviews with the RIN, LSE and the DCC to gather evidence on the value and impact of the project deliverables and outputs for the wider HE community

Outputs

The outputs from the project would be as follows

- 4-6 page case studies on each of the three project partners collated from the evidence gathered during the summative and formative interviews
- Evaluators Report (approx 10-12 pages) containing analysis of the cross-consortium challenges and successes, and the relevance of the outputs to the wider HE community

Outcomes

The evaluation would achieve the following outcomes:

- Each partner to have reflected on and recorded the experience from their perspective in such a way that it enables them to build future strategy and operations from the project
- An analysis and record of cross-partner issues, including challenges and successes, that serve to evaluate the 'consortium experience' and the value-added for working in this way
- An analysis of the relevance of the deliverables and outputs for the wider HE community, and as input for those providing support and training (RSP and DCC in this instance)

Timeline

September	Read project plan and project reports
October	Design and agree formative interview questions
November	Conduct formative telephone interviews
December	Short report on summative interviews
January/February	Design and agree summative interview questions and process
March	Conduct interviews with project partners
March	Conduct interviews with community stakeholders
April	Develop case studies and final report

This page last changed on Oct 27, 2008 by rice.

Advocacy and Promotion

Advocacy and promotion took a backseat to development over this period. The repository was declared "open for business" in a University newsletter in November. A few deposits have resulted, but an advocacy and promotion strategy will be developed and implemented over the next period to gain deposits.

The team worked with others in EDINA and Information Services at the University of Edinburgh and Les Carr at Southampton to host a successful "Repository Fringe" event during the summer, which attracted about 80 attendees from the UK and further afield. This was funded by the JISC Repositories and Preservation programme. The event was recorded and both powerpoints and streaming media are available via the website at <http://www.repositoryfringe.org>

Policy and Strategy

We understand that an open access requirement may be passing through the University Senate soon; this will effect ERA and there may be downstream repercussions/opportunities for Edinburgh DataShare. We have been discussing with the ERA managers how the two repositories will interoperate, and need to further discuss the linking of datasets with publications; especially as it is envisaged that full-text would only enter the ERA repository from the Publications Repository (originally used for the RAE), rather than directly into the DSpace repository.

We have been looking at the OpenDOAR tool for generating policies and working with Ann Green on the policy deliverable.

We have begun to make decisions about which formats in DSpace will be supported, e.g. preserved over the long-term and noting this DSpace.

We have written a Depositor Agreement and implemented it within DSpace.

We have implemented the option of the Public Dedication Domain and License for depositors to select and for end-users to receive.

Edinburgh DataShare has adopted the following procedure for converting proprietary database formats within the DSpace Institutional Repository. Further policies may follow regarding the preservation of databases in other formats (e.g. XML). In addition to creating and storing a preservation version, Edinburgh DataShare also intends to store the original proprietary version wherever practicable. The following outlines the procedure for converting an MS Access database using the Relational Database Management System SQL Server: SQL server loaded onto a local PC. The MS-Access database (Survey of Dedications to Saints in Medieval Scotland) is opened in SQL Server (ensuring that the table and column header names contain no spaces). A new database is created using the upsizing wizard and all the tables in the original database are selected and imported. The next step is to export and normalise the database schema by generating SQL Scripts. The tables are then exported (choose data source Microsoft OLE DB Provider for SQL Server) to a chosen destination as filetype Unicode. Some cleaning was required of the sql schema. The schema and the Unicode tables are then zipped and ingested into the repository along with the original version of the database.

Technical and Metadata Enhancements to IRs

MetaData Schema

The dublin core metadata fields to be used in Edinburgh DataShare repository were finalised and implemented in DSpace. This process included the identification of the fields to be used, which fields were mandatory, the control type to be used for the field and the wording of the textual hints for user guidance.

Embargo

A simple embargo feature was implemented in DSpace. One of the initial questions in the ingest process is whether the user would like to apply an embargo on the item. If the user elects for an embargo, a date field will appear in the first describe form where the user can specify the end date of the embargo. Until that date is reached the item bitstreams will not be available for download to non admin users, although the item metadata will still be visible and made available for harvesting.

License Options

The default DSpace ingest workflow was modified to accommodate an alternative end user license agreement. In the initial questions section of the ingest process the user can now select whether to have an Open Data Commons License or No License. If the Open Data Commons License option is chosen, the user will be asked to read and accept the license later in the process. If the user accepts the license, the license will be attached to the DSpace item. If the user chooses No License, the Rights metadata field will become mandatory. This has been implemented to be extensible due to the likelihood of additional licenses being added in the future.

Auto-account creation

The DSpace account creation process was deemed to be cumbersome, especially as it was possible to collect information about the visitor via the University's EASE authentication system and the Lightweight Directory Access Protocol (LDAP) server. From the visitor's University User Name it is possible to get the visitor's name and email address via and the university's LDAP server. With these details it is possible to automatically create a DSpace account. Should the user not have an LDAP entry, they are prompted for the details, as in the standard DSpace account creation process. DSpace was modified further to send a welcome and information email to the user once an account is created.

Automated Testing

Automated system tests have been written using HtmlUnit in an attempt to achieve high code coverage of extensions to DSpace. This should speed up and simplify regression testing in the future.

Spam Prevention

Added spam prevention measure to DSpace contact form.

Assessing Impact

We cannot comment on this as yet. The repository was only soft-launched at the beginning of November.

Oxford DataShare Progress Report - October 2008

Advocacy and Promotion

The *Scoping Digital Repository Services for Research Data Management* project is working in collaboration with the Oxford DataShare partners and has been involved in advocacy and promotion through a range of activities.

To find out about researchers' data management practices and main requirements for services, 36 interviews took place with researchers across disciplines and departments. The results highlighted researchers' need for support, tools and infrastructure to manage their data and comply with funders' requirements.

In addition, a Research Data Management Workshop was held in June and brought together researchers from Oxford and elsewhere to share their experiences with managing their data. The event also included talks from national initiatives, including Liz Lyon's report on dealing with data, the RIN report on data publication, the JISC-funded *Keeping Research Data Safe* study, and the Data Audit Framework.

The findings of these activities were published in July; the report *Findings of the Scoping Study and Research Data Management Workshop* can be found at: <http://www.ict.ox.ac.uk/odit/projects/digitalrepository/findings.xml>

These findings were also presented at the Repository Fringe conference in Edinburgh. In addition, the report published in July by the UK Research Data Service feasibility study made use of the findings of the report.

Currently, the project is conducting a consultation with service units across Oxford to find out about the data management services they offer. This consultation will be complemented by another workshop, in October, which will show examples of the data services provided and aims to raise discussion amongst service units in Oxford on best ways to support researchers.

The first part of the work for the Data Audit Framework has also started in Oxford. The Young Lives Project and a research group within the Oxford University Computing Laboratory have agreed to participate. Moreover, data resources currently published on the web are being added to the data inventory.

Policy and Strategy

Another ORA policy is nearing completion: the ORA services policy. The ORA content policy has been revised and updated. Neither includes data at this juncture because of the uncertainty of the scope of ORA for accepting data.

Technical and Metadata Enhancements to IRs

Oxford members of the DataShare project are in the process of depositing a collection of audio data files (currently in Microsoft WAVE (.wav) format) into the Oxford University Research Archive (ORA). The files were produced as part of an ESRC-funded research project: the researchers were required to deposit the data in the ESDS and

make it available as part of the condition of funding. However, the ESDS was not willing to accept these audio files.

The collection comprises several thousand files each about 20 seconds long, and a few dozen files that are perhaps 15 minutes long and total around 2.6Gb. Most are about 1 megabyte, but a few dozen perhaps are about 16 MB. The audio files form the data on which two published papers that are already in ORA were based. Copyright is owned by the University of Oxford and, as there are relatively few such data sets freely available, it is possible that they might be heavily accessed and used.

The collection of audio files comprises a large number of brief audio recordings of people reading short phrases. The recordings are of interest to linguists or phoneticians looking at details of pronunciation, and are likely to be more desirable as a group than individually. A serious researcher would probably want to download the whole corpus. To meet this need, the files will be added to ORA as separate items but also grouped together in a single corpus allowing a user to download all files. They will also be grouped in different sets determined by their original directory. Data abstracts will be used to describe the data.

A separate instance of Fedora will be set up to contain the data files and which will act as a proof of concept repository for data within the Oxford federated repositories: the scope of ORA for research data has not yet been settled. This 'DataBank' will be accessible from the existing ORA (i.e. items in ORA will reference the data files), but they will not initially be discoverable via a separate search facility. Each file, group of files and the whole collection will have its own jump off page displaying selected metadata with a link for access to the content. Each file will have a resolvable URL based on the Oxford UUID system with the main corpus having a Handle. The files will be described using DC terms and RDF XML will be used so they can easily be re-purposed and re-used.

Assessing Impact

The work of the Oxford DataShare partners has promoted awareness of the importance of actively managing and curating research data. Internally several groups have approached members of Oxford DataShare to enquire about how to best manage their data.

Also, many individuals from other institutions have expressed an interest in the project. We have had visits from staff from Monash University, Queensland University of Technology and the University of Sydney, as well as citations in the UKRDS report and QUT's *Practical Data Management: A Legal and Policy Guide*, see: <http://eprints.qut.edu.au/archive/00014923/>

**DataShare: Southampton Progress Report
October 2008**

Advocacy & Promotion

During this reporting period, the focus of advocacy work at Southampton has shifted towards the Data Audit, which is due to take place between December 08 and February 09. Over the summer, planning for the audit began and this included discussions with staff in the School of Social Sciences.

The Project Officer initially approached the *Deputy Head of School for Research* to gauge the level of interest. The Deputy Head's reaction was very positive and she made suggestions as to which research groups/departments might be audited. After subsequent meetings with the *Head of Gerontology* and the *Head of the Centre for Global Health, Population, Poverty & Policy*, it was agreed that the audit would cover the Division of Social Statistics and the Division of Gerontology.

These meetings were also used to discuss the audit forms provided by the DAF Methodology and have enabled the Project Officer to tailor the forms to suit the needs and types of research data used in the Divisions.

In September, Teresa McGowan was appointed as Auditor to work in partnership with the Project Officer. It was felt that a graduate from the Division of Social Statistics would have both the right skills and the staff contacts to make the audit a success. Teresa came recommended by the *Deputy Head of School for Research*. Since graduating with an MSc, Teresa has worked as a research assistant in the School and has recently led a series of focus groups as part of an ESRC funded research project. It is hoped that her contacts with School staff and understanding of social science research processes will be of considerable benefit to DataShare.

The other main area of work over the last 6 months has been the development of a metadata schema to describe datasets in ePrints Soton. It is intended that ePrints Soton will have a new *item type* entitled 'Dataset' and the metadata schema will be used to develop a new deposit form for datasets. Input from researchers has been sought on the selection of the most appropriate metadata elements and on the detail of the explanatory information to be provided on the deposit form.

Work to find suitable data for deposit has continued over this period and a potential deposit has been identified in the School of Geography. As yet, the Project Officer has been unable to make contact with the Professor concerned but this will be pursued.

Contact has been continued with the Professor of Film Studies who is developing a database of German-Speaking Emigres in British Cinema. With help from partners in Edinburgh, options for converting the PHP MySQL database to a preservation format are being explored.

Strategy and Policy

To date, the success of advocacy work to find datasets for deposit in ePrints Soton has been limited owing to issues discussed in the previous report. With the introduction of the data audit, the focus has moved away from issues of open access towards those of data management and data sharing within research groups. Research staff have already been found to be more receptive to these issues and it is expected that this new strategic direction will prove fruitful.

Technical & Metadata Enhancements to IRs

As mentioned above, the development of a metadata schema has taken place in this reporting period. The metadata schema is based on that developed for Edinburgh DataShare but it has been adapted in three ways. Firstly, as far as possible, it has been made consistent with the existing ePrints Soton repository. Secondly, it has been designed to work with the new ePrints 3.1 functionality. For example, ePrints 3.1 has a document versioning function that is far superior to that available previously and this has been utilised. Thirdly, it was necessary to take account of the fact that Edinburgh DataShare is a dedicated data repository, whereas ePrints Soton is an established institutional repository which is primarily populated with research papers.

In addition to discussions with School staff, the development of the metadata schema required lengthy consultation with the Repository Manager, a Developer at ePrints Services, the ePrints Soton Developer and Robin Rice.

The requirements of the metadata schema have now been incorporated into the upgrade schedule for ePrints Soton. Unfortunately, the expected completion date for the upgrade has slipped slightly but we hope to have a prototype to test in January 09. The schema may be modified when feedback is received from researchers in the new year.

Assessing Impact

See Policy and Strategy, above.

London School of Economics, Associate Partner Report

Data Librarian Post

LSE Library has been developing a new approach to the management and development of Data Library Services, following difficulties in recruiting to a single Data Library Manager post. We have identified the various strands of the service offered by the Library, and we are developing a team approach which will enable the flexible delivery and development of LSE Library services, along with the ability to actively participate in research and development activities in the field.

The Library has recently made the first appointment in this new team - with Dave Puplett taking up the post of Data Librarian on 6 October. Dave has worked in the Library since 2007, on a variety of projects focussing on improving access to electronic resources and exploiting new web technologies. He has recently completed work on the JISC funded VIF project dealing with version identification of academic research in digital repositories.

Dave will be the main contact for the Data Library service and he can be contacted on: 020 7955 7943 d.puplett@lse.ac.uk

Over the coming months we will continue to develop the Library's data services team and our commitment to the Data Library service remains as strong as ever.

Nicola Wright, Information Services Manager

NEEO Project

NEEO (Network of European Economists Online) is an EU-funded project which will address the lack of integration of academic output amongst premier European economics institutions by creating a powerful new research tool called Economists Online.

At UCL in September 2008 the NEEO project celebrated the completion of it's first year. To date all milestones and deliverables have been met and the first version of the portal is expected to be available in January 2009. Next month marks the beginning of work package 4 on datasets. The first meeting around datasets will be held during the NEEO meeting in Paris in January 2009. On the agenda will be a discussion of IPR, copyright and other issues with respect to datasets. By the end of the NEEO project each partner is expected to make 10 datasets available to create a total of 160 datasets.