



Project Document Cover Sheet

Project Information			
Project Acronym	DataShare		
Project Title	DISC-UK DataShare		
Start Date	March 2007	End Date	March 2009
Lead Institution	EDINA, University of Edinburgh		
Project Director	Peter Burnhill, Mark Brown		
Project Manager & contact details	Robin Rice, EDINA and Edinburgh University Data Library Main Library Bldg., George Square, Edinburgh EH8 9LJ R.Rice@ed.ac.uk , 0131 651 1431 or 0131 651 1317		
Partner Institutions	Universities of Edinburgh, Oxford, Southampton and London School of Economics		
Project Web URL	http://www.disc-uk.org/datashare.html		
Programme Name (and number)	<i>JISC Repositories and Preservation Programme: Repositories Start-up and Enhancement projects strand</i>		
Programme Manager	Andrew McGregor (Amber Thomas)		

Document Name			
Document Title	<i>DISC-UK DataShare Project Plan</i>		
Author(s) & project role	Robin Rice, Project Manager, and project team		
Date	28 September 2007	Filename	Project_Plan_DataShare.pdf
URL	http://www.disc-uk.org/deliverables.html		
Access	<input type="checkbox"/> Project and JISC internal		<input checked="" type="checkbox"/> General dissemination

Document History		
Version	Date	Comments
1	24 September 2007	



DISC-UK DataShare Project Plan



Overview of Project

1. Background

DISC-UK is an existing support network for data professionals working in universities (Data Information Specialists Committee-UK). DataShare is not an acronym, but is descriptive of the essence of the purpose of the project. We wish for the name to be meaningful to academics as well as librarians. It also has a synergy with an earlier successful JISC project in the 5/99 Learning & Teaching Programme, “Using Numeric Data in Learning and Teaching”, known as Datateach, which studied the barriers to the use of secondary data sources from the UKDA, MIMAS and EDINA, and covered issues around institutional support for learners and teachers in quantitative methods within the social sciences and related fields.

This project builds on an international network with a tradition of *data sharing and data archiving* dating back to the 1960s in the social sciences that has emerged more recently in other domains². By working together across four universities and internally with colleagues already engaged in managing open access repositories for e-Prints, this partnership will introduce and test a new model of data sharing and archiving to UK research institutions. By supporting academics within our institutions who wish to share datasets on which written research outputs are based, this network of institution-based data repositories develops a niche model for deposit of ‘orphaned datasets’ currently filled neither by centralised subject-domain data archives/centres/grids nor by e-Print-based institutional repositories (IRs).

For example, the social science survey of the StORe project found that nearly half of respondents had created some form of quantitative dataset (such as surveys), and over a third had produced a qualitative dataset (such as interviews). Yet 71% indicated they had never deposited data in a source repository such as the UK Data Archive.³ Similarly, the OpenDOAR registry of repositories indicates which repositories in the UK have a policy to accept datasets, but experience at the partners’ institutions shows that in actuality this is a long way from being the case. But the common policy to support only PDF filetypes undermines the ability to deposit datasets in forms that allow re-use in the sense of being able to manipulate the data for analysis in any way. We believe it is safe to say that active advocacy is not geared toward gaining deposit of research datasets in any existing UK university IR.⁴

¹ The project director and manager were the same people as in this project, and data librarians from two of the current partner institutions were involved (LSE and Edinburgh). See <http://datalib.ed.ac.uk/projects/datateach.html>

² Burnhill, P., Rice, R. and D. Geraci (2005). *The social side of science data sharing: distilling past efforts. Preserving Value Conference Proceedings*. Edinburgh: The Royal Society, 21-23 November 2005. Poster paper: <http://www.ukoln.ac.uk/events/pv-2005/pv-2005-final-poster-papers/040-poster-1.pdf> Poster: <http://www.ukoln.ac.uk/events/pv-2005/posters/burnhill-geraci-rice.pdf>

³ Burton, G. (2006). *The Source to Output Repository Project: Social Science*. [Unpublished.] <http://hdl.handle.net/1842/1418> (pp 13, 17).

⁴ Six out of 87 UK open access repositories listed in OpenDOAR record ‘datasets’ as a content type; aside from eCrystals at Southampton, DSpace@Cambridge and NDAD (National Digital Archive of Datasets) these returned only PDF articles in a search on ‘datasets’, and had no option to browse by dataset content type.

There are of course some notable centralised data archives and centres serving particular disciplines in the UK, such as the UK Data Archive/Economic and Social Data Service (UKDA/ESDS) for the social sciences and the NERC Data Centres for natural and environmental sciences. Other disciplines have created vast online databases on the Internet or over e-Research grid networks, which is the logical place for ‘publishing’ data outputs in those domains. (Illustrative case studies can be found in the *Joint Data Standards Study*, 2005.⁵) This project does not aim to challenge these existing institutions that have set internationally recognised high standards in data archiving, management and curation, nor the model of centralised data repositories by subject area. It does however, wish to explore the role of filling in the gaps left open by the paucity of coverage of dedicated data archives, and in doing so, gain leverage from being able to work closely and directly with potential depositors at one’s own institution. Indeed, the lifecycle approach to data sharing encourages intervention at the earliest stages of a research project to ensure adequate consent, documentation etc., are achieved for the data to be usable by others.⁶

From Open Access to Open Data

Open access repositories are still a relatively new feature in scholarly communication. The Open Archives Initiative was only launched in 1999, for example⁷; many of the thorny problems associated with the management of institutional repositories have yet to be solved (including both technical and cultural). Nevertheless libraries and funding agencies in North America have already begun to address the arguably new role of research libraries in the curation of data. (For example, at the NSF/ARL “Workshop on New Collaborative Relationships: The Role of Academic Libraries in the Digital Data Universe” in September 2006⁸ and the Science Commons Workshop in October 2006.⁹) Similar issues were explored in an RIN workshop in London on 5 December, 2006.¹⁰

“Open Data” as a concept is quickly gathering momentum as the latest term in the “open” trilogy along with Open Source (for software) and Open Access (for research outputs, primarily e-Prints).¹¹ It indicates a recognition that there is a rising level of expectation among users for complete access to an intellectual work, not only the final published post-print, but the body of evidence drawn on to create that final output. This is compatible with the scientific method of allowing replication of results by others, and the rich tradition of secondary analysis in the social sciences and other population-based research domains. It is also in line with several recent initiatives to open up publicly-funded research data to public availability.¹²

⁵ The Digital Archiving Consultancy, The Bioinformatics Research Centre (University of Glasgow) and The National e-Science Centre (NeSC) (2005). *Large-scale data sharing in the life sciences: Data standards, incentives, barriers and funding models (the Joint data standards study)*. <http://www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC002552>

⁶ Humphrey, C.K., Estabrooks, C.A., Norris, J.R., Smith, J.E. and K.L. Hesketh (2000). *Archivist on board: Contributions to the research team*. **Forum Qualitative Sozialforschung / Forum: Qualitative Social Research** [On-line Journal] 1(3). <http://qualitative-research.net/fqs/fqs-eng.htm>

⁷ Suber, P. (2006). *Timeline of the Open Access Movement*. **Open Access News** <http://www.earlham.edu/~peters/fos/timeline.htm>

⁸ <http://www.arl.org/info/events/nsfworkshop.html>

⁹ “Creating a vision for making scientific data accessible across disciplines” <http://www.spatial.maine.edu/icfs/>

¹⁰ “Getting the most out of data, Making the most of research” <http://www.rin.ac.uk/digital-data-workshop>

¹¹ For example, a wikipedia entry for the term was added only as recently as October 2006.

¹² For just one example see the 2004 “OECD Ministerial Declaration on access to research data from public funding” (Annex 1) http://www.oecd.org/document/0,2340,en_2649_34487_25998799_1_1_1_1,00.html. The new Medical Research Council (MRC) data sharing requirements are based on the OECD principles.

However, there are reasons data are not always provided in a completely open way. Aside from commercial value, these may include confidentiality concerns, copyright questions, and sheer complexity. Many scholars are more comfortable with an informal method of sharing, so they can assess the use to which the data will be put and decide whether to give the requestor access on a case by case basis.¹³ This is one of the reasons datasets have not been incorporated into IRs to date; the project will need to address this.

2. Aims and Objectives

The project's **overall aim** is 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.

Objectives:

1. Build capacity of institutional repositories in UKHE to respond to the unique requirements of research datasets as a new 'document type'.
2. Use a range of open source software repository solutions – Eprints, DSpace, Fedora – to provide exemplars and add-on tools for managing datasets as institutional repository items.
3. Produce and disseminate findings – in cooperation with the Repositories Support Project (RSP) and the Repositories Research Team (RRT) - to inform library and repository managers about the organisational and technical issues associated with the deposit of research data.
4. Work with the RSP, Digital Curation Centre (DCC) and others to identify training needs and solutions for increasing skills of information professionals in UKHE for managing research data.

3. Overall Approach

The existing DISC-UK partnership – involving staff at the Universities of Oxford, Edinburgh, Southampton and the London School of Economics—will form the foundation for the project's activity. Each of these universities has a strong commitment to their IRs and are active in enhancing them. DISC-UK members will work closely with staff involved with IR management and development at their own institutions to pilot successful models for incorporating deposit of research data into institutional repositories. In this way, the partner institutions will be early adopters and provide exemplars, particularly but not exclusively for numeric datasets in population-based research areas—this is the domain expertise that DISC-UK partners have in common.

The DataShare project is based on a distributed model in which each participating partner is responsible for the work on their own repositories, yet experience, support and knowledge are shared in order to increase levels of success. This builds on the existing informal collaboration of DISC-UK members for improving their data libraries and models of data support at each institution. It will also bring academic data libraries in closer contact with e-Prints repository managers and develop new forms of cooperation between these distinct groups of information professionals within academic environments. The advantage for the broader community is to provide exemplars for a range of approaches and policies in which to embed the deposit and stewardship of datasets in IRs. These will be demonstrated using the three main repository solutions in the UK: EPrints, DSpace and Fedora.

This project plan is intended to deliver outcomes beneficial to the wider JISC community, as well as to take into account the particular situations at each of the participating institutions.

¹³ Both the StORe and GRADE projects have gathered evidence about this.

4. Project Outputs

WP1: Project Management

- Project plan (this document with appendices)
- Bi-annual progress reports and final report to JISC
- Information-rich website geared toward various audiences
- Internal meeting notes or virtual recordings from online meetings
- Shared filestore and email list archive for project staff
- Presentations from meetings with external invitees
- Correspondence (liaison) with RSP, RRT, DCC, in addition to data centres, research councils and open source repository communities, and academic data communities
- Increased skills of project staff through professional development opportunities
- Signed consortium agreement
- Evaluation report

WP2: Institutional Repository Development

- State of the Art Review
- 'Backdoor' data surveys/ partial audits at each institution (info gathering exercise to identify potential depositors, may or may not include actual survey of staff or interview questions)
- Local publicity (e.g. newsletter articles, flyers)
- Policy documents (defining scope of data repository; copyright policies; depositor agreement forms, etc)
- Depositor guidelines
- Job specifications for new staff
- Reports on TRAC or DRAMBORA methodology used by each site to measure progress toward trusted repository status (facilitated by consultant)
- Develop repository workflow model for deposit of exemplar datasets

WP3: Technology

- Customisation code for DSpace, EPrints, Fedora – shared with OSS communities
- Deliverable on linking eprints with datasets in context of individual IR (DSpace)
- Deliverable on use of DDI metadata schema within individual IR (EPrints)
- Deliverable on Shibboleth access to data in individual IR (Eprints)
- Deliverable on a data content model (Fedora)
- Deliverable on interoperability with UKDA
- Deliverable on documented exemplars for Web 2.0 data sites
- Technology watch bookmarks and RSS feeds on Bluedot social bookmarking site

WP4: Outreach and dissemination

- Dissemination plan based on stakeholder analysis and experience in first year
- Experience-based outputs for repository managers, e.g. use cases, scenarios or toolkits
- Papers and presentations at relevant national and international conferences
- Documentation, best practice guidelines, fact sheets on range of topics
- Design a training programme for support of research data analysis, management, sharing in libraries/institutions

5. Project Outcomes

- Exemplars of the process, pitfalls and successful outcomes of setting up an institutional data repository service at each of the four institutions
- 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
- Enhancements to partners' IRs including testing for trusted repository status
- Technical watch on e-Research, VREs, Web 2.0 and related developments
- Papers, presentations and online dissemination of collected knowledge

Outputs from the project will help to demystify data as complex objects in repositories, and assist other IRs in overcoming barriers to incorporating research data. By building on lessons learnt from recent JISC-funded data repository projects such as CLADDIER and GRADE¹⁴ the project will help realise the vision of the Digital Repositories Review of a “coherent aggregation of content from a network of institutional repositories”,¹⁵ and more particularly of the Digital Repositories Roadmap, see for example the milestone under Data: “Institutions need to invest in research data repositories”.¹⁶

6. Stakeholder Analysis

Stakeholder	Interest / stake	Importance
JISC	Funder; promoter of IRs	High
Open access / open data advocates (worldwide)	Promoting their goals	Medium
Librarians & repository managers (mainly UK)	Learn from our project about dealing with research data	Medium
Research councils and funders	Help to make funded data accessible (or compete with a centralised vision?)	Medium
Relevant JISC projects and services	Deliverables may further their goals (or compete?)	Medium
Open Source Software communities	Enhancements / reputation to their products	Low
IASSIST and soc sci data community (worldwide)	Watching outcomes, may influence future work	Low
Publishers	May offer a way for publishers to mandate making data available (or compete?)	Low
Partner institutions: library & info service mgmt	Employer of staff and investment in project	High
Partner institutions: library & info service staff	Peers; potential promoters of our service to academic staff	Low
Partner institutions: senior faculty & management	Potential champions	Medium
Partner institutions: academic staff & soc sci departments who are early adopters	Make their research available openly; quality of service	High
DISC-UK members	Success of project as sponsor; recognition of group	High

7. Risk Analysis

Risk	Probability (1-5)	Severity (1-5)	Score (P x S)	Action to Prevent/Manage Risk
Staffing				
Delay in start; recruitment problems	3	5	15	Identified project officers, liaison roles, and project manager are

¹⁴ The data cluster of the Digital Repository Programme projects can be found on the Repository Wiki, http://www.ukoln.ac.uk/repositories/digirep/index/Category:Data_repositories_cluster .

¹⁵ Heery, R. and Anderson, S. (2005). *Digital repositories review*. http://www.jisc.ac.uk/uploaded_documents/digital-repositories-review-2005.pdf

¹⁶ Heery, R. and Powell, A. (2006). *Digital repositories roadmap: Looking forward*. <http://www.ukoln.ac.uk/repositories/publications/roadmap-200604/>

				already in post. Work can be done even if recruiting on technical posts causes delay.
Risk	Probability (1-5)	Severity (1-5)	Score (P x S)	Action to Prevent/Manage Risk
Loss of key staff before end of project.	2	4	8	Ensure embeddedness of project in institutional practices; involve both data library and repository staff in project setup and workflows so that knowledge never rests with one person only. Keep project management transparent to project team.
Organisational				
Difficulties in developing a consortium agreement and robust partnership	1	4	4	DISC-UK is an active partnership already. As part of bid preparation, foundation for cooperation between data libraries and repository managers has been laid.
Partner institution withdraws support for project or fails to embed in future services	2	4	8	Integrate activities with existing IR structure as much as possible; use local governance structures. Communicate 'lessons learned' including the good and bad to broader community.
Technical				
Technical solutions not available for identified problems.	3	3	9	Include technical staff in project who can develop open source software and common tools; contribute solutions to broader community to reduce technical barriers to acceptance of datasets by other institutions. Communicate appropriateness of IR for data to user communities; not same as data publishing; manage expectations. Maintain technical watch and communication with other projects. If necessary, migrate to one of the other systems (e.g. e-Prints, DSpace, Fedora).
Legal and External				
Lack of acceptance of institutional repositories as place for depositing/managing datasets..	4	5	20	Be active in repository programme to show how DISC-UK model fits into landscape; communicate with key stakeholders including data centres and research councils; put sufficient effort into outreach with academics at each partner institution; understand their needs; recommend deposit in large data centres such as UKDA whenever appropriate.
Illegal content in repository leads to lawsuit against partner institution or individual	1	5	5	Ensure appropriate policies, guidelines and procedures are written and accessible. Implement a swift take-down policy in response to complaints. Educate users about preparing public use datasets and copyright/

				confidentiality/consent issues. Train project staff where possible.
--	--	--	--	--

8. Standards

Name of standard or specification	Version	Notes
XHTML		For project website
OAI-PMH		For harvesting
Dublin Core		For metadata within repositories
XML		For data formats
Data Documentation Initiative (DDI)		For fully marked up social science quantitative dataset exemplars including codebook (variable level) information
XSLT		For web publishing of XML datasets

9. Technical Development

There are a number of difficult issues that must in some way be resolved for any IR to accept responsibility for deposit of datasets in earnest, including ingest, metadata schemas, storage, file format migration, unique identifiers & levels of granularity, applying OAI-PMH to datasets, and making content available for centralised harvesting by data type. Since the DataShare project will in this sense be a first true attempt, gaining understanding from international precedents will be crucial. Contact will be made with particular IR managers known to have experience in managing research datasets, such as at Johns Hopkins University and MIT in the US. A consultant well-versed in these developments in the US—Ann Green from Digital Life Cycle Research & Consulting, New Haven, Connecticut, USA¹⁷—will assist each partner in aligning its development plans with institutional strategies and situations, taking into account: policies, procedures, software choices, workflow, organisational structure, suitable levels of support (e.g. for assisted deposit and data curation), preservation planning and scope. Ms Green will visit each institution early in the project and facilitate meetings of data and repository staff, advising on these areas and training needs of participating staff. Ms Green will revisit the partnership in the middle of the project to assist the sites to test their repositories against the DCC DRAMBORA toolkit and/or the CRL Trustworthy Repositories Audit & Certification: Criteria and Checklist, as well as towards the end to participate in the dissemination strategy.

Each partner institution will directly allocate staff time to the project (up to 25% per individual), and in return will gain additional effort for technical support (50% at each institution except Oxford, at 25%). Oxford and Southampton will use the additional effort for data curation and advocacy; Edinburgh and LSE will hire or buy out additional technical staff to design and support the repository software itself. Servers (hardware and software) do not need to be purchased as they already exist at each site, though one or two staff machines are included in the budget.

Edinburgh will use DSpace, Oxford will use Fedora, and both LSE and Southampton will use EPrints, all open source repository software used in UKHE. Software enhancements needed for datasets will be identified in the course of the project and carried out by technical staff at the institutions.

10. Intellectual Property Rights

¹⁷ Green, A. and Gutmann M. (2007). *Building partnerships among Social Science researchers, institution-based repositories, and domain specific data archives*. OCLC Systems and Services: International Digital Library Perspectives. 23(1). Preprint: <http://hdl.handle.net/2027.42/41214>

For written outputs, copyright will belong to the authors or their institutions, but will be licensed to JISC in perpetuity, and will also be made available open access within the author's institutional repositories and on the DISC-UK website. Academic publication of some articles will be pursued. Software will be developed within open source environments. There are IPR issues to do with deposit of research data which will be explored as part of the project, with each institution coming up with suitable policies, procedures and guidelines for users.

Project Resources

11. Project Partners

EDINA & Edinburgh University Data Library:

Edinburgh Research Archive (ERA) - <http://www.era.lib.ed.ac.uk/>

Edinburgh University Data Library - <http://datalib.ed.ac.uk/> *EDINA* – <http://edina.ac.uk/>

Main contact: Robin Rice, R.Rice@ed.ac.uk

London School of Economics:

Main contact: Luis Martinez, L.Martinez@lse.ed.ac.uk

LSE Research Online - <http://eprints.lse.ac.uk/>

London School of Economics Data Library - <http://www.lse.ac.uk/library/datlib/Default.htm>

LSE Research Laboratory Data Service -<http://rlab.lse.ac.uk/data/>

University of Oxford (Nuffield College):

Main contact: Jane Roberts, jane.roberts@nuffield.ox.ac.uk

Oxford Research Archive - <http://ora.ouls.ox.ac.uk/>

Oxford Data Library (Nuffield College) - <http://www.nuff.ox.ac.uk/projects/datalibrary/>

University of Southampton:

Main contact: Harry Gibbs, T.A.Gibbs@SOTON.AC.UK

e-Prints Soton - <http://eprints.soton.ac.uk/>

University of Southampton Library Data Resources -

www.soton.ac.uk/library/resources/collections/data/

Digital Life Cycle Research & Consulting

Contact: Ann Green, green.ann@gmail.com

External Evaluator

To be determined.

The consortium agreement is yet to be finalised and signed by the partner institutions.

12. Project Management

Framework

The Project Directors will oversee the project's direction as a whole, with the Project Manager, who will bring strategic decisions to their attention. As this project is largely about embedding new developments within institutional repository services, it is most important that the partner institutions establish "buy-in" and set up their own governance for project activities as part of their existing services. This will take different forms, but some form of 'steering committee' at each institution will be used to report and monitor progress; in some cases, this already exists, as at Oxford, where repository activity is formally governed by such a faculty committee.

The external evaluation will cover both the project's overall success and impact as well as the achievement of embedding project deliverables within the partner institutions.

The project manager will spend 40% of her time on project management and an additional 15% on the Edinburgh repository development.

Communication within the project is achieved through a closed project listserv, a shared filestore, one to one email and telephone communication, occasional face to face meetings, and a regular (roughly monthly) telecon meeting (this is hoped to be replaced soon by Agora, an online meeting tool that includes video of each participant). Participants at the telecon meeting include the DISC-UK project officers and the repository managers; it is chaired by the project manager. The meetings will ensure consensus for decisions about the project as a whole and act as a locus for quality assurance of deliverables.

Project staff report to their line managers at their institutions. Line managers are responsible for ensuring training needs are met, but the project manager will identify opportunities where appropriate and approve project funds to be spent on training.

Project Team:

Project Directors

- Peter Burnhill - EDINA National Data Centre, University of Edinburgh
- Mark Brown - University of Southampton

Project Manager

- Robin Rice - EDINA National Data Centre and Edinburgh University Data Library

R.Rice@ed.ac.uk

+44 (0)131 651 1431 or +44 (0)131 651 1317

EDINA

160 Causewayside

Edinburgh EH9 1PR

Project Officers

- Harry Gibbs - University of Southampton Library
- Stuart Macdonald - EDINA National Data Centre and Edinburgh University Data Library
- Luis Martinez - London School of Economics Data Library
- Tanvi Desai - LSE Research Laboratory Data Service
- Jane Roberts - Oxford Data Library (Nuffield College)

Repository Managers

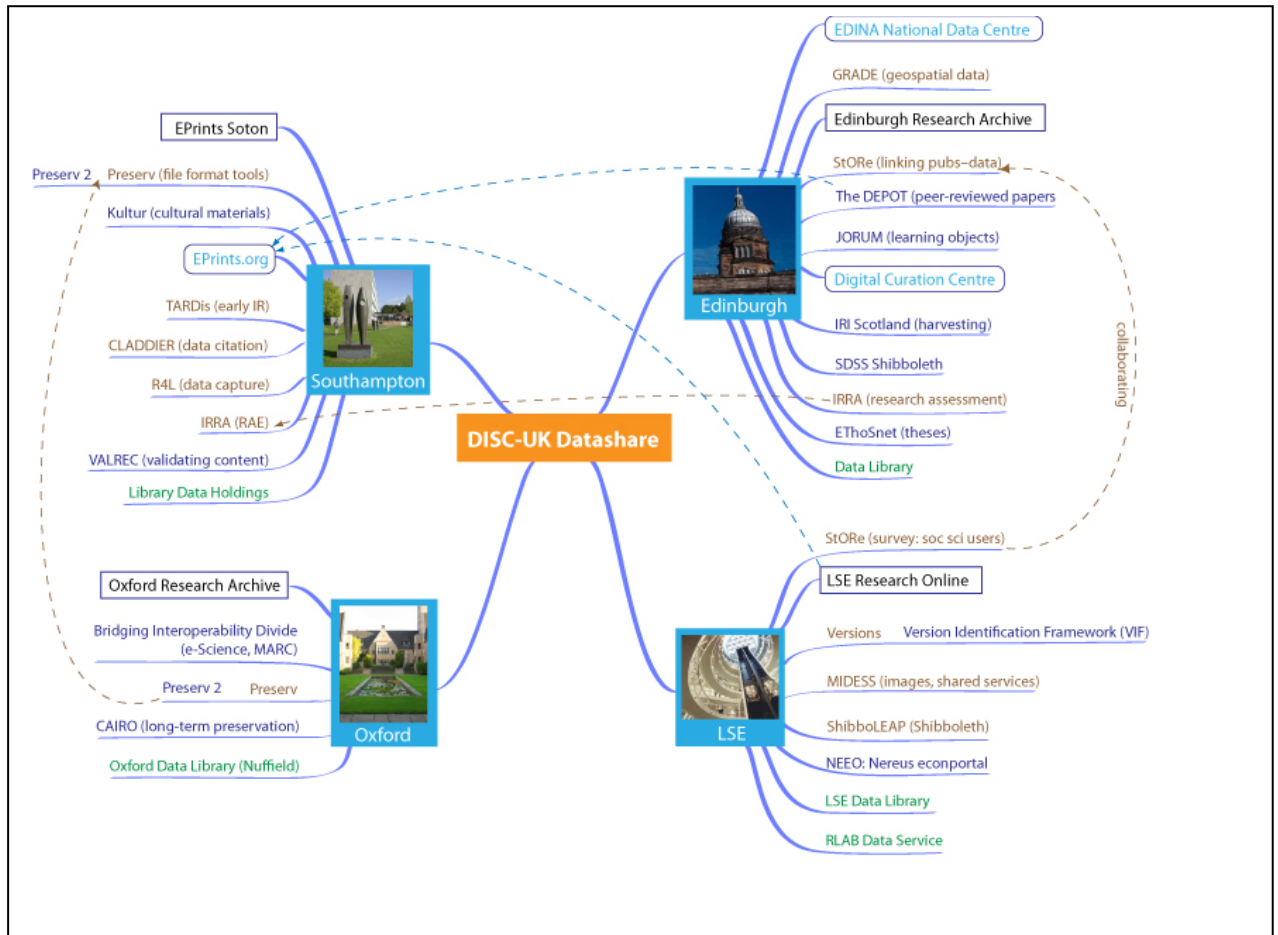
- John MacColl - Digital Library Division, University of Edinburgh
- Theo Andrew - Digital Library Division, University of Edinburgh
- Frances Shipsey - Information Services, LSE Library
- Sally Rumsey - Oxford University Research Archive
- Neil Jefferies - IT Development & Strategy Team, Oxford University Library Services
- Wendy White - University of Southampton Libraries

Project Consultant

- Ann Green - Digital Life Cycle Research & Consulting, Newhaven, Conn., USA

External Evaluator and technical staff to be determined.

Repository Experience of Project Partner Institutions



- Legend for colours on diagram:
- green: local data services
 - dark blue: current projects
 - brown: past projects
 - light blue: services
 - black: partner institutional repositories
 - brown dotted line: collaboration between services
 - blue dotted line: EPrints.org software user

This diagram shows the extensive repository experience of the four project partner institutions which the project can draw upon.

Keywords: file format tools, cultural materials, early IR, data citation, data capture, RAE, validating content, e-Science, MARC, long-term preservation, geospatial data, linking pubs-data, peer reviewed papers, learning objects, harvesting, research assessment, theses, survey: soc sci users, versioning, images, shared services, Shibboleth, econportal

For clickable diagram with hyperlinks, see project website at <http://www.disc-uk.org/team.html#diagram>

13. Programme Support

The project would benefit from clarification with research councils about the role of IRs in making funded data available, if that is something the JISC programme can do. Discussion has already taken place regarding liaison with other services on training and dissemination, and possible external evaluators. Policies and procedures from other repositories will also be useful. Support for digital preservation planning would also be appreciated, especially with regard to the variety of data formats.

14. Budget

See Appendix A. (No changes from proposal.)

Detailed Project Planning

15. Workpackages

See Appendix B.

16. Evaluation Plan

Because this is a new area for UK IRs, an external evaluator familiar with UKHE and the JISC information environment will be chosen early in the project to help create an evaluation plan, covering both progress made at each institution, and the impact of the project as a whole on the broader community. Both summative and formative forms of evaluation shall be used.

17. Quality Plan

Output Timing	Quality criteria	QA method(s)	Evidence of compliance	Quality responsibilities	Quality tools (if applicable)
Project Management deliverables (WP1)	Success of project viewed internally and externally	Evaluation process; monthly project meetings; feedback from external contacts; timely delivery of outputs	Evaluation report; correspondence with programme manager, project team, improved skills of team members, evidence in progress reports and project website	Project manager, programme manager, project directors, local steering committees and project team, evaluator	Evaluation tools as determined by external evaluator with project manager
Institutional repository development (WP2)	Interface / metadata suitable for datasets; written policies in place; promotional activities and materials; interaction with	Consultant to monitor progress at each IR and work through trusted repository criteria (either TRAC or DRAMBORA)	IR interface; policy documents; promotional documents; meeting activity reported in progress reports; consultant	Consultant, local managers and steering committees, consultant, project team	DRAMBORA and TRAC

	researchers; data content in IRs; evidence of preservation planning	with each institution. Project team to report progress to project manager for JISC reports.	reports		
Output	Quality criteria	QA method(s)	Evidence of compliance	Quality responsibilities	Quality tools (if applicable)
Technology (WP3)	Clean and robust code; seamless additions; usable interface	Feedback from open source communities, institutional line management practices, user logs and usability testing with users	Demonstrations, access to pilot systems, written code, documentation, use cases, project website	Line managers and local steering committees, project manager, evaluator	Validation tools for some code/mark-up languages
Outreach and dissemination (WP4)	Well written, concise dissemination materials, geared toward intended audiences, useful new information, communicating experience	Peer review (where appropriate), designated proofreader for project outputs, team to read and comment before wider circulation, outputs approved by line manager & project manager, work with established services where appropriate (RSP, DCC)	Feedback accommodated from readers before 'publication', acceptance in academic conferences & journals, user feedback, downloads from website, feedback sought from other projects	Proofreader, line manager, project manager, project team	Use of Harvard style citations in reports.

18. Dissemination Plan

Timing	Dissemination Activity	Audience	Purpose	Key Message
Throughout	University newsletters, lunchtime seminars, small group discussions	Local academic staff; postgraduates	Raising awareness and engage researchers	Benefits of depositing data in IR

June 07	Presentation at CURL/SCONUL Task Force on e-Research	Librarians	Introduce project	Role of data librarians in supporting e-research
September 07	Poster session at European Conference on Digital Libraries	Librarians and computer scientists	Showcase EDINA repository project	How EDINA is building institutional capacity for supporting community generated content
October 07	Presentation at the International Internet Librarian	Information professionals	Raising awareness about <i>open data</i>	Linking open data to open software, open access and web 2.0
November 07	Presentation on data deposit to AHDS workshop on self-archiving (invited speaker)	AHDS user community	Contribution to AHDS exit strategy; clarify role of IRs in accepting data	Not yet determined
August 07 to End	Project website, social bookmarks, blog	Everyone	Explain project, raise awareness, disseminate outputs, engage communities	About DISC-UK and DataShare
December 07	Poster Session at DCC Conference (if accepted)	Digital curators	Show institutional role for curating data	IRs have a role to play in the data lifecycle
April 08	Southampton Repository Conference (possible paper)	Repository managers	To show how repositories can accommodate data	DataShare partners will provide exemplars of best practice for community
May 08	IASSIST (possible paper)	Social Science data professionals	Show experience in assisting researchers manage their data	Evolution of data librarians' role in helping users

Further items in the dissemination plan will be elaborated for each progress report, particularly for year two.

19. Exit and Sustainability Plans

Project Outputs	Action for Take-up & Embedding	Action for Exit
Published papers	Deposited in IR	Preservation
Website and dissemination materials	Maintain website	Web archiving
Advocacy for data sharing	DISC-UK staff continue in daily work	Those who are permanent staff only
Training materials	Training providers (such as RSP, DCC, national data centres) to take up	Responsibility for conducting training negotiated with service providers
Technology	Implemented in IRs, documented, 'published' with open source community	Documentation accessible to new staff, open source community

The funding will greatly assist the partners to get 'up and running' with an ambitious plan. There is a momentum to this project which both precedes and will last beyond the life of the project. Two elements will help to ensure the sustainability of the project's outputs for the partners involved: the DISC-UK network itself and its connections with an international community of data professionals; and the commitment of each partner institution to the goals of the project and to enhancing their repositories for incorporation of research data. The project will work with both the RSP and the DCC to make its external outputs available and useful beyond the life of the project.

Appendices

Appendix A. Project Budget

Appendix B. Workpackages



JISC Project Plan Budget

Directly Incurred Staff	Mar-07	April 07-March 08	April 08-March 09	Total £
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]
Total Directly Incurred Staff	6,743	79,735	84,387	170,865
Non Staff				
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]
Total Directly Incurred Non Staff	3,020	18,324	16,436	37,780
Directly Incurred Total	9,763	98,059	100,823	208,645
Directly Allocated				
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED] ²
Directly Allocated Total	4,998	59,792	60,156	124,947
Indirect Costs	8,497	101,960	93,463	203,920
Total Project Costs	23,258	259,812	254,442	537,512
Amount Requested from JISC	12,227	127,628	127,927	267,782

Institutional Contributions	11,031	132,184	126,515	269,730
Percentage contributions over the life of the project	JISC	50	Partners	50



JISC WORK PACKAGE

WORKPACKAGES	Month	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
1: Project Management																										
2: Institutional Repository Development																										
3: Technology																										
4: Outreach and Dissemination																										

Project start date: *1 March 2007*

Project completion date: *27 February 2009*

Duration: *24 months*

Workpackage and activity	Earliest start date	Latest completion date	Outputs (clearly indicate deliverables & reports in bold)	Milestone	Responsibility
--------------------------	---------------------	------------------------	--	-----------	----------------

				Milestone	Responsibility
WORKPACKAGE 1: Project Management					
<i>Objective: to ensure efficient and timely delivery of project outcomes.</i>	01/03/07	28/02/2009			
1. Project planning and leadership			Produce Project plan ; make adjustments to plan as needed; monitor progress and responsibilities. Signed Consortium Agreement .		Project Manager, RR (Edinburgh); Project Directors, PB, MB (Edinburgh, Southampton)
2. Facilitate collaboration and communication among partners			Project meetings scheduled; collaboration software set up; ensure engagement of all partners; assist project planning at partner sites as needed.		RR (Edin)
3. Contact and reporting with JISC			6-monthly progress reports and final report ; liaison with programme manager and other project managers; attendance at programme meetings; participate in online discussion.		RR (Edin)
4. Financial administration			Maintain budget; administer and monitor flow of funds to partners; ensure efficient and timely expenditure of funds.		RR & IE (Edin)
5. Liaison between consultant and partners			Develop consultant activity (e.g. TRAC or DRAMBORA self-audits); monitor progress; arrange and facilitate	Kick-off meetings at each	RR (Edin); AG (Digital Lifecycle)

Workpackage and activity	Earliest start date	Latest completion date	Outputs (clearly indicate deliverables & reports in bold)	Milestone	Responsibility
			meetings.	site (April 2007)	Consulting)
6. Establish and maintain project page, website			EDINA project page with key information; project-wide website with current awareness information about data and institutional repositories		RR, SM, JC (Edin); all
7. Ensure professional development opportunities for project staff			Data librarians to learn about repositories; repository managers to learn about data; technical staff to undertake training as needed		RR (Edin); AG (DLC); all
8. Oversee internal and external evaluation			Arrange for both formative and summative evaluation covering both progress made at each institution, and the impact of the project as a whole on the broader community. Commission an external evaluator familiar with UKHE, research data, and the JISC IE.		RR (Edin); external evaluator
WORKPACKAGE 2: Institutional Repository Development	01/05/07	28/02/2009			
<u>Objective:</u> <i>to use a range of open source software repository solutions - Eprints, DSpace, Fedora - to provide exemplars and add-on tools for managing datasets as institutional repository items. Develop each partner's IR to accept deposit of datasets; inform each other about best practices.</i>					
9. State of the Art Review to inform our set-up activities				State of the Art Review publicly	HG (Southampton); all

Workpackage and activity	Earliest start date	Latest completion date	Outputs (clearly indicate deliverables & reports in bold)	Milestone	Responsibility
				released	
10. Data Audit (backdoor info gathering, e.g through Research Office, and surveys)			Internal information for repository staff		All
11. Prepare user scenarios? list of benefits; legal briefing; practice on 'friendly' researchers			User scenarios, etc. on project website		All
12. Local advocacy & finding early adopters; publicity			Newsletter articles, powerpoints, etc.		All
13. Define scope & policies (e.g. file formats, metadata, preservation)			Policy documents on IR websites		All
14. Develop depositor agreement (copyright etc.)			Depositor agreement integrated into IR deposit interface		All
15. Write/compile guidance for depositors; e.g. data management guidelines; data sharing policies from research councils; creating anonymised public use files			Policy documents and guidance for depositors on IR websites		All
16. Spec new jobs & hire staff				New staff in place	Edinburgh; LSE; Oxford
17. Implement EPrints, DSpace, and Fedora systems (data IRs) at each of the sites (either new or existing repositories).			Repository system in place for development. Usability testing of repositories.		All
18. Customise IRs for ingest interface for datasets			Includes determining mandatory fields		All
19. Customise IRs for metadata for datasets			Includes exploration of simple DC & qualified DC		All
20. Customise IRs for preservation of datasets			Use of Pronom DROID tool or other file format registry & checksums for data validation.		All
21. Customise IRs for OAI compliance of datasets			IR's data contents are harvested by major repository search engines	Register IR in OpenDoar	All
22. Work with project consultant to measure progress towards trusted data repository			Each institution will use RLG TRAC or Drambora methodology to identify		AG; All

Workpackage and activity	Earliest start date	Latest completion date	Outputs (clearly indicate deliverables & reports in bold)	Milestone	Responsibility
			priorities for preservation planning ??		
WORKPACKAGE 3: Technology	01/05/2007	28/02/2009			
Objective: <i>Build capacity of institutional repositories to respond to the unique requirements of research datasets as a new 'document type'.</i>					
23. Customisation to link to publications in IR			Any new code will be made available to OSS communities		LSE; Edinburgh
24. Customise EPrints to use DDI as metadata schema			Includes OAI compliance		LSE
25. Implement a plug-in for Shibboleth access and authentication to resources			Includes both authentication for deposit and exploration of restricted access to datasets by Virtual Research Organisations (user-defined groups)		LSE
26. Develop a Fedora data model			DDI and (?) other XML schemas (?)		Oxford
27. Explore interoperability with the UKDA			Goal is to allow access to datasets derived from UKDA datasets through seamless access interface to UKDA registration. Need to negotiate with UKDA.		DISC-UK
28. Exemplars for Web 2.0 data sites (e.g. swivel.com, many eyes), mash-ups (e.g. user's own data with google maps) and custom-built websites (e.g. using SVG for choropleth maps)			Depends on early adopters' and their needs.		RR & SM (Edin)
29. Technology Watch: e-Research & Repositories			Bookmarks on DataShare bluedot site; other written outputs; possible tasks for technology implementation		LM (LSE)
30. Technology Watch: Web 2.0 & Repositories			Bookmarks on DataShare bluedot site; other written outputs; possible tasks for		SM (Edin)

Workpackage and activity	Earliest start date	Latest completion date	Outputs (clearly indicate deliverables & reports in bold)	Milestone	Responsibility
			technology implementation		
31. Technology Watch: Preservation & Repositories			Bookmarks on DataShare bluedot site; other written outputs; possible tasks for technology implementation		RR (Edin)
WORKPACKAGE 4: Outreach and Dissemination					
<i>Objective: to produce and disseminate findings to inform library and repository managers about the organisational and technical issues associated with the deposit of research data.</i>	01/05/2007	28/02/2009			
32. Develop a dissemination plan for the second year with emphasis on deliverables and based on a complete stakeholder analysis.			Original dissemination plan will continue to grow as part of progress reports.		RR
33. Cooperate with the RSP and DCC to customise experience-based outputs for dissemination to UKHE community, including tool-kits for data management.			Meetings and correspondence with RSP and DCC.		Various
34. Papers will be submitted to relevant conferences and journals by DISC-UK partners. Presentations will be given by request to other meetings.			As identified, to be included on project website.		Various
35. Experience-based outputs including use cases, documentation, best practice guidelines, and fact sheets will be written by the project team on a wide range of topics, in consultation with RSP and others as to desired topics.			As identified, to be included on project website.		Various
36. A training programme will be designed (but not delivered within the project) to teach			Correspondence and meetings with key collaborators; progress towards training		DISC-UK members with

Workpackage and activity	Earliest start date	Latest completion date	Outputs (clearly indicate deliverables & reports in bold)	Milestone	Responsibility
skills to information professionals who wish to support staff and students in their use of research data in their institutions for analysis, management and sharing in repositories.			requirements identified.		collaborators as identified

Members of Project Team:

Project Directors

- Peter Burnhill - EDINA National Data Centre, University of Edinburgh
- Mark Brown - University of Southampton

Project Manager

- Robin Rice - EDINA National Data Centre and Edinburgh University Data Library

Project Officers

- Harry Gibbs - University of Southampton Library
- Stuart Macdonald - EDINA National Data Centre and Edinburgh University Data Library
- Luis Martinez - London School of Economics Data Library
- Tanvi Desai - LSE Research Laboratory Data Service
- Jane Roberts - Oxford Data Library (Nuffield College)

Repository Managers

- John MacColl - Digital Library Division, University of Edinburgh
- Theo Andrew - Digital Library Division, University of Edinburgh
- Frances Shipsey - Information Services, LSE Library
- Sally Rumsey - Oxford University Research Archive
- Neil Jefferies - IT Development & Strategy Team, Oxford University Library Services
- Wendy White - University of Southampton Libraries

Project Consultant

- Ann Green - Digital Life Cycle Research & Consulting, Newhaven, Conn., USA

Technical officers tba