JISC

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)	*JISC Repositories and Preservation Programme: Repositories Start-up and Enhancement projects strand*		
Programme Manager	Andrew McGregor (Amber Thomas)		

Document Name			
Document Title	*DISC-UK DataShare Progress Report, October 2007*		
Author(s) & project role	Robin Rice, Project Manager, and project team		
Date	26 October 2007	**Filename**	DataShare_Progress_Report_Oct07.pdf
URL			
Access	■ Project and JISC internal	□ General dissemination	

Document History		
Version	**Date**	**Comments**
1	26 October 2007	

JISC Progress Report October 2007

Overview of Project

Grant Statement

No changes. For clarification, the following dates are noted.

23 Nov, 2006: Bid submitted.

23 Feb. 2007: Feedback received on bid with invitation to revise and re-submit.

14 Mar, 2007: Informal permission to ramp up project activity through spending funds given by JISC.

6 Apr, 2007: Revised bid submitted.

27 Apr, 2007: Grant letter approving funding of project received.

1 May, 2007: Clock starts ticking in terms of grant letter obligations and deadlines.

2. Aims and Objectives
No changes.

3. Overall Approach
No changes. As this is a distributed project, please see Appendix B for Institutional summaries of progress on enhancements to IRs during the period.

4. Project Outputs
Summarise progress during the reporting period and milestones/deliverables achieved.

The project got off to a strong start with the kick-off week meetings with Ann Green, our consultant, during the last week of April. Internal momentum was continued through monthly teleconference meetings, the project listserv, and where opportunities to meet face to face arose. The final project plan was submitted at the end of September. A number of dissemination opportunities have occurred, as well as professional development for the project team. Regarding developments to each repository, the lateness of the funding letter and the sudden project start meant for some that time was needed to clarify how new posts would fit, write job specifications, and carry out hiring processes. For others where hiring was not an issue, the aims of DataShare needed to be prioritised along with other existing plans to IR development. Nevertheless, the project is now in full swing at each partner site, and beginning to make itself heard in the broader community. The team continues to work well together and we look forward to welcoming new staff on to the project. Please see the Workpackages section (heading 15) for a detailed list of outputs from the reporting period.

5. Project Outcomes

We are aware that we are staking out new territory, where others have mainly only dabbled. The State-of-the-Art Review was an essential piece of research on which to base our activities (we do not want to be re-inventing any wheels), and has revealed many of the environmental issues we operate

within.The AHRC's decision not to renew funding for the AHDS has been a major event during the period and has focussed both positive and negative opinions about the ability of IRs to rise to the occasion of preserving complex data. Liz Lyon's report on Dealing with Data was another important milestone for the community and has provided a number of recommendations at both a policy and institutional level. The DataShare continuum graphic has received positive feedback from contacts, and has helped us to show the boundaries of what the project aims to cover, in an area that has many players and heterogeneous approaches, not to mention disciplinary differences. It also helps to show where open data or open access fit in, which is still an alien concept to many. Finally, where the AHRC seems to be heading for decentralisation, the ESRC seems to be headed for consolidation and centralisation, as our talks with UKDA staff have revealed. Perhaps this is an area where the JISC programme can help clarify and build the relationships between research councils, universities operating IRs, and data providers such as AHDS and ESDS.

6. Stakeholder Analysis

We continue to try to build relationships with local academics to provide us with the valuable experience that only handling actual research data and projects will give us. However, we need to rehearse and perfect ways of approaching academics that will appeal to their immediate needs rather than fill in the missing pieces of our project goals. The feedback documented in the Southampton local report shows that we will not likely win over a majority of academics within the project period. Instead, the strategy will be on seeking out the willing, and providing workable exemplars with different data types.

The project itself has been a useful mechanism for making contacts with like-minded support staff at other institutions and has been attracting positive attention at dissemination events and through email correspondence with the project manager. (See Workpackages, Item 15, for list of contacts made, both within the UK and internationally.)

7. Risk Analysis

- *Delay in start; recruitment problems*
  As anticipated, the delayed funding letter plus immediate start caused delays in putting new staff in post. Edinburgh has successfully appointed; LSE is currently advertising.
- *Loss of key staff before end of project*
  We have recently learned that two members of the team will be leaving their posts: John MacColl, Director, Digital Library Group, at Edinburgh, and Luis Martinez, Data Librarian (and DISC-UK member), at LSE. It is hoped that Luis will be able to participate at least tangentially in the project in his new post at Oxford from January. Edinburgh plans are as yet unknown; LSE intends to recruit a replacement. Luis Martinez' post at LSE is likely to be filled by someone without previous data librarian experience, which could be a setback for the project. On the other hand, there will be additional support on the technical side from the FAR project at LSE.

8. Standards

No change.

9. Technical Development

No change.

10. Intellectual Property Rights

No change.

Project Resources

11. Project Partners

No change.

12. Project Management

Staffing

As noted in 7) Risk Analysis, two team members will be leaving, plus one (Theo Andrew, Repository Manager, ERA, Edinburgh) has been seconded 50% to another post and will need to be partially replaced. Only Luis Martinez' post has designated FTE effort on the project.

Edinburgh has advertised and successfully recruited a Software Engineer / DSpace Developer for its technical input to the project. See Edinburgh progress report in Appendix B for job description. George Hamilton will be joining EDINA for the project on 26 November.

LSE is currently advertising an Institutional Repository Programmer, closing date 2 November. See Appendix B for job description.

13. Programme Support

Our project is sufficiently different from most other projects in the programme that we have not developed close ties. However, we have agreed to stay in touch with Sam Peploe's OJIMS project since the last programme meeting, as we were the only two involved with data. The project manager looks forward to meeting other projects interested in "research data" from the newer rounds. In addition we have contacts with most of the "data cluster" projects from the Digital Repositories programme (finished). We are also aware of other synergies, e.g. complex objects, and maintain contact through the jisc-rpprog-sue list and jisc-repositories list. We have been encouraged to participate in the RSP wiki by the programme manager and will do so. As included in the project plan, each partner is well-connected with repository work, as shown in the diagram of repository experience at http://www.disc-uk.org/diagram.html. As mentioned in Project Outcomes, above, the project may

require a "boost" from the programme or other external force in terms of advocacy for open data to be successful reaching local academics.

14. Budget

See Appendix A. The budget report template provided by JISC includes directly allocated expenditure, which is being spent according to the original budget, but we do not have the detailed costings on this to fill in. Also, we have added a column for committed funds to show what each partner plans to spend but has not yet invoiced EDINA for. In some cases this is due to late staffing starts, but will be made up for with additional FTE effort.

Detailed Project Planning

15. Workpackages

WP1: Project Management

- *Project plan*

  <u>Sept, 2007:</u> **Project plan** submitted and accepted by JISC programme manager, A McGregor. (Workpackages submitted in August and revised following feedback.) Placed on website minus appendices.

- *Bi-annual progress reports and reporting to JISC*

  <u>Oct, 2007:</u> This document is first **progress report**.

  <u>Ongoing</u> communication with programme manager by email as needed.

  <u>July, 2007:</u> **Programme meeting** in London, attended by Robin Rice (both days), Peter Burnhill, and Tanvi Desai (one day each).

- *Information-rich website geared toward various audiences*

  <u>May-June, 2007:</u> EDINA project page and JISC page completed.

  <u>Sept, 2007:</u> DataShare logo and new style for DISC-UK **website** developed by EDINA's graphic designer in consultation with project manager. DISC-UK website and project web information integrated and launched.

  <u>Ongoing:</u> Project information updated on website by Edinburgh project team.

  <u>Ongoing:</u> Blue dot **social bookmarking** site used to track related developments in the UK and globally. Tagging used to allow future creation of subsets or bibliographies.

- *Internal meeting notes or virtual recordings from online meetings*

  <u>15 Mar:</u> Robin Rice and Peter Burnhill meet Ann Green by telecon to plan kick-off week.

  <u>23-27 April:</u> **Kick-off meetings** organised by project manager and hosted at each partner site with accompanying presentations and brainstorming activities led by project consultant, Ann Green. Presentations and Excel file of combined brainstorming activities made available to team.

  <u>16 May:</u> DISC-UK members Robin Rice, Stuart Macdonald, Jane Roberts, Tanvi Desai, Luis Martinez, and Harry Gibbs met with Ann Green during IASSIST conference in Montreal.

<u>14 June:</u> DISC-UK members Robin Rice, Stuart Macdonald, Luis Martinez, and Harry Gibbs have project meeting in Edinburgh before SCONUL workshop.

<u>28 Jun, 26 July, 26 Sept, 25 Oct:</u> Project team monthly **telecon meetings** took place, chaired by Robin Rice. Open University's Meet-o-matic used to schedule meetings. Future meetings to be held using **Agora** online video-conferencing (free) software developed at Leeds Uni, installed on an EDINA server. (Installation progressing gradually.) Functionality will include sound and video using webcams at client PCs, shared whiteboard, recordings of meetings.

- *Shared filestore and email list archive for project staff*

- *Presentations from meetings with external invitees*

<u>3 April:</u> "DISCUK-DataShare" private **JISCmail list** and filestore set up by Luis Martinez and Robin Rice for project team. Presentations from kick-off meetings available there, including those by Ann Green, Anne Robertson (EDINA), Andrew McHugh (DCC), Jessie Hey (Southampton), Midess (LSE), and notes from Sally Rumsey's demo of ORA (Oxford University Research Archive).

- *Correspondence (liaison) with RSP, RRT, DCC, in addition to data centres, research councils and open source repository communities, and academic data communities*

<u>Ongoing:</u> **Liaison** with **RSP, RRT and DCC** staff has begun, partly via email and partly through interactions at meetings including JISC Programme meeting, RIN workshop, OAI5 conference, e-Scidir Workshop, CASPAR/DPE meeting, ECDL conference. The project manager has discussed the project with Bill Hubbard (RSP), Mary Robinson (RSP), Mahendra Mahey (RRT), Liz Lyon (UKOLN and DCC), Maureen Pennock (RSP and DCC), Seamus Ross, Andrew McHugh, Joy Davidson, Perla Innocenti (DCC at Glasgow), Chris Rusbridge, Graham Pryor (DCC at Edinburgh), as well as Sheila Anderson (AHDS), and Sam Peploe (BADC). Luis Martinez at LSE and John MacColl at Edinburgh have discussed it with Stéphane Goldstein (RIN) and reported to the SCONUL Task Force on e-Research about the project. Patricia Killiard, Cambridge University Library, kindly shared DSpace Developer job spec with Edinburgh. Elizabeth Bishop, Qualidata, has been in touch with the project manager regarding the ESRC funded Timescapes Project based at Leeds. Alison Macdonald and Philip Lord, The Digital Archiving Consultancy Limited, London, invited Robin Rice and Peter Burnhill to the e-Sci Dir workshop in Lisbon to consider policy recommendations on scientific digital repositories for the EU (both attended). Stuart Macdonald accepted invitation to participate in user evaluation of StORe pilot by NCESS at Manchester, 29 July.

International contacts established or revived in light of the project by the project manager include: Myron Gutmann, Director of ICPSR, Michigan; Lu Chou, Data and Information Services Center, University of Wisconsin-Madison; Maude Frances, The University of New South Wales, Australia; MacKenzie Smith, Associate Director for Technology, MIT Libraries, Boston (regarding DSpace PLEDGE project); Frank Barjak, AVROSS coordinator, Switzerland; Tim Dilauro, Digital Knowledge Centre, Sheridan Libraries Johns Hopkins University, Washington, DC; Lee Dirks, Director, Scholarly Communications, Microsoft, Redmond, Washington; Prof. Christine Borgman, Department of Information Studies,

Graduate School of Education and Information Science, University of California, Los Angeles. Luis Martinez initiated contact with Belinda Weaver, Manager, UQ eSpace, The University of Queensland Library.

27 Mar: DISC-UK members (Tanvi Desai, Luis Martinez, Jane Roberts, Harry Gibbs, Stuart Macdonald) meet with **UK Data Archive/ESDS** Director, Kevin Schurer and Karen Dennison at their invitation to discuss common interests and improvements to ESDS services. Hosted at LSE. A member of DISC-UK was invited to join the ESDS **steering committee**. Tanvi Desai is serving the first term.

20 July: Robin Rice sent **'DISC-UK Response to RIN consultation'** by email. Available on project website, http://www.disc-uk.org/docs/DISCUKresponsetoRINconsultation.pdf. See http://www.rin.ac.uk/data-principles for RIN draft, *Stewardship of digital research data - principles and guidelines,* for which comments were being sought.

2 Aug: Meeting in Edinburgh, about **StORe phase 2** plans and DataShare, with Ken Miller, UKDA, and project team members Robin Rice, Stuart Macdonald and Luis Martinez.

12 Sep: Second meeting between DISC-UK members (TD, LM, JR, HG, SM, RR) and UKDA/ESDS Director, Kevin Schurer, and Karen Dennison held at LSE. DataShare project discussed as part of agenda. Change of ESRC/UKDA policy toward accepting all datasets discussed, in light of plans for their pilot Fedora repository (StORe 2). Other relevant topics discussed included changes to licensing government data due to the Statistics Act, EU funding of CESSDA network of national Data Archives, and aspirations for training of data archival professionals, as well as a possible panel or paper for next year's IASSIST conference.

Oct: Wendy White asked to join committee of the **National Centre for Research Methods** to "contribute to the development of ESRC strategy in the general field of sustaining on-line resources."

- *Increased skills of project staff through professional development opportunities*

  11 April: Stuart Macdonald attended SHERPA roadshow, St. Andrews.

  18-20 April: Robin Rice attended OAI 5 workshop at CERN, Switzerland (and gave a poster session for the Depot, a different EDINA project).

  14-18 May: Harry Gibbs attended **IASSIST** workshop and conference, held in Montreal, for first time, through project funding. Other DISC-UK members attended through other funding. (IASSIST is the International Association for Social Science Information and Services Technology and is the main professional organisation for data librarians. The theme was *Building Global Knowledge Communities with Open Data.)*

  5-6 June: Several members of project team attended **JISC Digital Repositories Conference** (some with other 'hats' on).

  14-15 June: Harry Gibbs and Robin Rice attended the SCONUL workshop hosted by NeSC, Edinburgh, *Support for e-Research: Filling the Library Skills Gap*, at which Stuart Macdonald and Luis Martinez gave a presentation about DISC-UK and announced the new project (see WP 4).

<u>24-26 Oct:</u> Luis Martinez attending 3-day workshop on "Using DDI 3.0 to Support Preservation, Management, Access, and Dissemination Systems for Social Science Data" Wadern, Germany.

<u>29 Oct:</u> Tanvi Desai attending "What's Happening to British Statistics" informational event at the Royal Statistical Society and reporting back (about the Statistics Act and repercussions).
• Signed consortium agreement

- *Evaluation report*

  Both the consortium agreement and the selection of the external evaluator have yet to be done. In progress.

WP2: Institutional Repository Development

- *State-of-the-Art Review*

  The **State-of-the-Art Review** has been researched, written and completed by Harry Gibbs, quality assured by Wendy White (line manager) and Robin Rice (project manager) and proof-read by Jane Roberts. It is on the website in PDF format at http://www.disc-uk.org/docs/state-of-the-art-review.pdf, linked from the project deliverables page, as well as the DISC-UK Publications and Presentations page, and deposited in e-Prints Soton, http://eprints.soton.ac.uk/48596/. It has been announced on the following email lists and blogs: JISC-RPPROG-SUE@JISCMAIL.AC.UK, JISC-Repositories@jiscmail.ac.uk, iasst-l@lists.columbia.edu (for IASSIST members) and the IASSIST Communiqué (public blog), http://iassistblog.org/?p=78, and SPARC-OpenData@arl.org. Clifford Lynch, director CNI, forwarded it to the US-based list, CNI-announce@cni.org, with the following preface:

  > "This is a very helpful review of the current state of data sharing practices from a UK perspective that has come out as part of the JISC funded DISC-UK DataShare project. I'm fowarding the full announcement from the SPARC Open Data list."

- *'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)*

  See partner site reports in Appendix B.

- *Local publicity (e.g. newsletter articles, flyers)*

  See partner site reports in Appendix B.

- *Policy documents (defining scope of data repository; copyright policies; depositor agreement forms, etc)*

- *Depositor guidelines*

  Not yet written or revised for datasets.

- *Job specifications for new staff*

  Both Edinburgh and LSE have advertised for a technical **repository developer post**. The job descriptions are not appended but available on request. So far, they would not differ greatly from other IR technical post job descriptions.

- *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*
  This work not yet done..

WP3: Technology

- *Customisation code for DSpace, EPrints, Fedora – shared with OSS communities*
  IRs not yet customised for datasets.
- *Deliverable on linking eprints with datasets in context of individual IR (DSpace)*
  Dependent on building data repository first; technical post to begin 26 November.
- *Deliverable on use of DDI metadata schema within individual IR (EPrints)*
  An initial report about adapting the **DDI** is expected by end of the year.
- *Deliverable on Shibboleth access to data in individual IR (Eprints)*
  Progress has been made in terms of communication among staff at LSE about using Shibboleth for privileged access to data, as in virtual organisations. Work is to be combined with effort of the **FAR project**, contact is Simon McLeish:
  https://gabriel.lse.ac.uk/twiki/bin/view/Projects/FAR/WebHome.
- *Deliverable on a data content model (Fedora)*
  No progress to report.
- *Deliverable on interoperability with UKDA*
  Communications with UKDA make this seem unlikely, as UKDA have used the StORe project to develop their own pilot fedora repository, in which they hope to collect datasets directly from users at institutions, whether their projects are funded by ESRC or not. They do not envisage cooperation with institutional repositories in terms of interoperability or harvesting. DISC-UK and UKDA have agreed to keep each other informed of progress.
- *Deliverable on documented exemplars for Web 2.0 data sites*
  A draft briefing paper on **Web 2.0 data sites** has been written by Stuart Macdonald and will be published on the website after completion and quality assurance.
- *Technology watch bookmarks and RSS feeds on Bluedot social bookmarking site*
  Since 1 May, **60 bookmarks** have been added, tagged, and annotated, to the Bluedot site at http://bluedot.us/users/DataShare.

WP4: Outreach and dissemination

- *Dissemination plan based on stakeholder analysis and experience in first year*
  Dissemination plan for year 1 is in the project plan. This will be updated at the next progress report.
- *Experience-based outputs for repository managers, e.g. use cases, scenarios or toolkits*
  Nothing has been written yet.
- *Papers and presentations at relevant national and international conferences*
  Macdonald, Stuart and Luis Martinez. "e-Research and the Data Librarian" (presented at the CURL/SCONUL Task Force organised workshop Support for e-Research: Filling the

Library Skills Gap, e-Science Institute, Edinburgh, UK, June 14-15, 2007).
http://www.disc-uk.org/outreach/NeSC_Jun_07.ppt

Rice, Robin, Burnhill, Peter, Rees, Christine, and Anne Robertson. "Repository Junction and
Beyond at the EDINA (UK) National Data Centre" *Research and Advanced Technology
for Digital Libraries: 11th European Conference, ECDL 2007 Proceedings* (September
2007) Budapest, Hungary: Springer (pp 543-5) [Poster paper]. Published version:
http://www.springerlink.com/content/82746560l1337771/fulltext.pdf [publisher's
website]. Actual poster in PDF presented by Robin Rice, 18 September, 2007:
http://www.disc-uk.org/docs/ECDLposter.pdf. (Poster covered four EDINA repository
projects and services – Jorum, the Depot, GRADE and DISC-UK DataShare.)

Martinez, Luis and Stuart Macdonald. "Open the Data Doors to Perception," presented at the
Internet Librarian International 2007: *New Realities, Roles, Resources*, London, UK,
October 8-9, 2007. http://www.disc-uk.org/outreach/IIL_noteless.ppt

Green, Ann. "Connecting the Dots among digital repositories, data services, and social
science researchers," presentation to the bi-annual meeting of the Official
Representatives (ORs) of ICPSR (Inter-university Consortium for Political and Social
Research), University of Michigan, Ann Arbor. http://www.icpsr.umich.edu/ICPSR/

- *Documentation, best practice guidelines, fact sheets on range of topics*
  <u>Sept 2007</u>: **Data Sharing Continuum graphic**, by Robin Rice [PDF] http://www.disc-uk.org/docs/data_sharing_continuum.pdf created to show how the project goals fit into the wide range of practices for curating research data.
- *Design a training programme for support of research data analysis, management, sharing in libraries/institutions*
  Not due until the end of the project, but UKDA is potentially interested in this area and we are in communication with them.

16. Evaluation Plan

External evaluators are currently being considered.

17. Quality Assurance Plan

The QA plan worked well for recent deliverables, including the State-of-the-Art Review.

18. Dissemination Plan
No change, as project plan was completed quite recently.

19. Exit/Sustainability Plan

On track as planned.

Appendixes

Appendix A. Project Budget

Appendix B. Partner Site Progress Reports

JISC Programme Management Framework
22 December 2003