



Project Document Cover Sheet

Project Information			
Project Acronym	EIDCSR		
Project Title	Embedding Institutional Data Curation Services in Research		
Start Date	01 April 2009	End Date	30 September 2010
Lead Institution	University of Oxford, Oxford University Computing Services (OUCS)		
Project Director	Dr Michael Fraser		
Project Manager & contact details	Luis Martinez-Uribe OUCS / Oxford e-Research Centre 7 Keble Road Oxford OX1 3QG luis.martinez-uribe@oerc.ox.ac.uk 01865610607		
Partner Institutions			
Project Web URL	http://projects.oucs.ox.ac.uk/eidcsr/		
Programme Name (and number)	JISC Information Environment 2009-11 (12/08)		
Programme Manager	Neil Grindley		

Document Name			
Document Title	Project Plan		
Reporting Period			
Author(s) & project role	Luis Martinez-Uribe, Project Manager		
Date	28 July 2009	Filename	EIDCSR_ProjectPlan_v2.0.doc
URL			
Access	<input checked="" type="checkbox"/> Project and JISC internal		<input type="checkbox"/> General dissemination

Document History		
Version	Date	Comments
V 1.0	16-Apr-09	Project Manager's first draft
V 1.1	1-May-09	Revision of Executive Board and Project Working Group; addition of IBM workpackage; modification of stakeholder analysis.
V 2.0	28 July 2009	Revision of WP 3, WP 6 and WP 7.



JISC Project Plan: Embedding Institutional Data Curation Services in Research

Overview of Project

1. Background

“There is a growing international understanding of the value of preserving research data and the necessity of developing an infrastructure to support this. ... not all research data within these disciplines [those that have UK research data centres] will be housed there and there are also many other disciplines and datasets which need to be maintained within institutions” (Beagrie et al. 2008)¹

The Embedding Institutional Data Curation Services in Research (EIDCSR) project arises from the range of activities carried out through an Oxford scoping study², a cross-agency collaborative effort to scope the requirements to manage and curate research data generated by Oxford researchers. A requirements gathering exercise took place with around 40 interviews with researchers across disciplines. Service units in Oxford were also consulted to identify their “data management services” and identify gaps in service provision. Moreover, the project contributed to the UK Research Data Service feasibility study and the JISC funded DISC-UK DataShare project.

2. Aims and Objectives

The project is an institutional collaboration bringing together the Cardio Mechano-Electric Feedback Group (CMEFG), the Computational Biology Group (CBG), the Library (OULS) and Computing services (OUCS), the Oxford e-Research Centre (OeRC), the Office of the Director of IT (ODIT) and the Research Service Office (RSO). In addition to this, IBM will contribute to the project.

Aim

The aim of the project is to address the data preservation requirements of two collaborating research groups in Oxford, by scoping the curation and preservation requirements for research data and embedding selected elements of the digital curation lifecycle, including policy, workflow, and sustainability solutions within the research process at an early stage. The workflows generated by the project are intended to scale to include other research domains and the outputs should be of use to other research intensive institutions.

Objectives

1. Embedding institutional services for data curation for two specific research communities that can then be expanded to other research disciplines;
2. Introduce curatorial practices within research workflows to add value to everyday scholarly work;
3. Develop and coordinate institutional and service level policies and economic models for the management, preservation and sharing of research data;
4. Investigate the roles and responsibilities of service providers in Oxford to support their researchers with the management and curation of research data by developing a deep

¹ Beagrie, N., Chruszcz, J., Lavoie, B. (2008) Keeping research data safe:
<http://www.jisc.ac.uk/publications/publications/keepingresearchdatasafe.aspx>

² The scoping digital repository services for research data management project:
<http://www.ict.ox.ac.uk/odit/projects/digitalrepository/>

understanding of research workflows with data and how they may interface with institutional services.

3. Overall Approach

The overall approach follows a three-stage process:

Stage 1. Analysis: The project will undertake a need analysis for undertaking data preservation work using the Data Audit Framework (DAF) Methodology. This work will build on previous DAF work in Oxford and will complete gaps in the requirements gathering and auditing of data assets. This stage will include the identification of metadata standards to describe and preserve these data.

Stage 2. Pre-implementation: During the pre-implementation phase, the project will translate the identified requirements into draft policy and business models, and technical implementation plans. The latter will focus on the development of embedded workflow modules that interface with existing services. This phase will also include reaching an agreement on metadata standards to describe and preserve the data from the participating research groups.

Stage 3. Implementation: The implementation phase will develop the policy framework, workflow and sustainability models by integrating existing services to address the requirements identified in the first stage. The three main services to interface are the Hierarchical File Server (HFS), the Fedora Digital Assets Management System (DAMS) and the local storage infrastructure.

There is clearly a symbiotic relationship between the development of data management policies and plans, the provision of technical infrastructure to support their implementation, and the application of a sustainability model for the overall service. The project plan allows for an iterative approach to the development of all three, both at the central and local levels.

4. Project Outputs

- A report detailing the requirements from the two research groups as well their data holdings and management practices, based on the DAF methodology;
- A recommended set of metadata standards relating to the description, administration and preservation of research data generated by the participating research groups;
- Software to provide a digital curation workflow module integrating the research data lifecycle with institutional services like the HFS and Fedora DAMS;
- Report on the strategic provision of federated data storage and access infrastructure within a devolved environment for research data;
- Annotated draft institutional research data management policy framework;
- Report on the application of costing and sustainability frameworks developed by JISC-funded projects;
- Project website, blog and RSS-supported bookmarks of relevant activities and publications;
- Two workshop with reports together with publications and presentations in relevant journals and conferences;
- A final report, describing the process of implementing the exemplar preservation solution and making recommendations on how JISC should consider continuing work in this area.

5. Project Outcomes

- Curation and preservation service definitions together with tools to help curate research output from two internationally known research groups in Oxford, offering a useful real-life case study for other institutions grappling with similar issues.
- A better understanding of viable models, including economic, for addressing the curation and preservation of research data from the moment of creation.
- An improved awareness about the benefits of actively managing and preserving research data in HE institutions.
- Examples on how to move from data audit to institutional service for the management and curation of research data.

- An extension of previous JISC work providing disciplinary data case studies demonstrating the benefits of multidisciplinary research data sharing and cost models for research data preservation.
- The development of services that have the potential to be expanded to other research domains (within and beyond the University of Oxford), providing sustainable research data preservation capabilities.

6. Stakeholder Analysis

Stakeholder	Interest / stake	Importance
University of Oxford – research groups and service units (OULS, OUCS, OeRC and Research Services Office)	The University of Oxford is committed to the project; the participating research groups are large-scale research activities with well-defined preservation needs; the service units have defined, interlocking roles and responsibilities for research support.	High
JISC	The curation of research data is a strategic priority for both the Integrated Information Environment and the Support of Research sub-committees.	High
UCISA/RUGIT	UCISA is the association representing HE in the provision of information systems. RUGIT comprises the IT directors and similar from the Russell Group of Universities. Both groups have at various times stated the importance of research data curation to their members.	High
The Digital Curation Centre	The DCC is a centre of expertise on digital curation, is engaging with staff in HEI with an interest in data management through the Research Data Management Forum and is involved in providing data curation training through the DCC 101 course.	High
The Research Information Network	Research data is one of the main themes for RIN activities and they have published several influential reports on data management and data sharing.	High
UK data centres	UK data centres such as the Economic and Social Data Service, the NERC data centres or the Archaeology Data Service are centres of excellence in the management and preservation of research data; and Oxford will engage with these through the project.	Medium
Research funding agencies	RCUK, and e.g. The Wellcome Trust, are developing requirements relating to data management and data sharing for funded activities; all of the research councils as well as other funding bodies intend to make preservation of data a condition of funding.	Medium
UK Research Data Service	The University of Oxford is working with UKRDS consortium to plan Pathfinder services and expects to form one of the main university Pathfinder centres in 2009/2010.	Medium
Other relevant projects and services	Other projects investigating issues related to the management and curation of research data may provide useful tools and information and may use some of the outputs of this project.	Medium

7. Risk Analysis

Risk	Probability (1-5)	Severity (1-5)	Score (P x S)	Action to Prevent/Manage Risk
Staffing				
Failure to recruit appropriate staff for the project.	3	4	12	Redeployment of existing staff will be investigated as well as timely recruitment of new staff.
Loss of key staff before end of the project.	2	3	6	Embed project in institutional practices, ensuring a number of individuals have expertise and willingness to assume responsibilities.
Organisational				
Expectations mismatch between project and research groups.	2	3	6	Ensure continued engagement via Project Working Group and Executive Board to maintain initial buy-in and momentum.
Lack of coordination between project stakeholders.	1	3	3	Ensure clear reporting and communication lines; take advantage of existing institutional communication structures.
Technical				
Development of workflow integration with technical systems proves to be more complex.	2	3	6	Technical implementation planning will provide early warning and potential to adapt; learn from similar projects.
Lack of domain metadata standards for data.	2	2	4	Information professionals and domain experts will be consulted from an early stage in the project.

8. Standards

The project website will comply with all accessibility and mark-up standards. The metadata required to describe, administer and preserve the research data will comply with research community standards as well as those available at the JISC Standards Catalogue. The rest of the technical development work required as part of the digital curation workflow will be based on service oriented architecture standards.

9. Technical Development

The project will implement technical components that will be largely developed within a service-oriented architecture. A workflow module will be developed around the HFS, using the Tivoli API, to enable the early capture of metadata; identification of the data provider and curator roles; and allow seamless upload and retrieval of data from the HFS. It should be noted that the HFS is a tape-based storage system and therefore, the workflow module is intended for long-term storage rather than general filestore facilities. In addition to this, the project will take advantage of Web Service components developed as part of the BID Project (and more recently the BR11 Project) to integrate the metadata packaging and management services of the Fedora DAMS within the workflow module.

10. Intellectual Property Rights

Any IPR resulting from this project will remain the property of the organisation generating it. Under the University of Oxford's policy on intellectual property (which covers all University employees and students), the University claims ownership of a range of intellectual property rights with commercial potential. The University does not assert any claim to the ownership of copyright in artistic works, books, articles or lectures, apart from those specifically commissioned by the University. Results arising from projects funded by the JISC at Oxford would therefore usually be owned in the first instance by the University as the employing institution. The University seeks to maximise the commercial potential of its intellectual property through its wholly-owned technology transfer company, ISIS Innovation Ltd. In accordance with the desires of the Information Environment and e-Research Programme, however, it is proposed to release project deliverables under either a Creative Commons license or, in the case of software, under an open source software license to maximize the benefit for the wider community.

Project Resources

11. Project Partners

The EIDCSR project is based wholly at the University of Oxford.

12. Project Management

An Executive Board comprising internal and external representatives from the key stakeholder communities will support the work of the project. This group will be responsible for helping to embed the project activities in the broader community by raising awareness of the existence of it and communicating expertise into the project. The group will also be expected to adopt a directing role meeting three times for the duration of the project. There will also be a reporting line to the Oxford Digital Repositories Steering Group.

Executive Board	
Dr Michael Fraser	PI
Prof. Paul W. Jeffreys	Co – Investigator
Prof. David Gavaghan	Co – Investigator
Prof. Peter Kohl	Co – Investigator
Neil Grindley	JISC Programme Manager
Luis Martinez-Urbe	Project Manager and Secretary to the Board
	OULS representative
	RSO representative
	OeRC representative
	IBM representative (by invitation)
	RIN representative
	RUGIT representative
	DCC representative

The Project Working Group comprising members of the project team and others from the research groups and service units in Oxford will facilitate an agile an iterative approach to the development of quality project outputs and to ensure that the project meets the aims and objectives

Project Working Group		
Dr Michael Fraser	PI	mike.fraser@oucs.ox.ac.uk Tel: 01865 283 343
Prof. Paul W. Jeffreys	Co – Investigator	paul.jeffreys@odit.ox.ac.uk Tel: 01865 273229
Prof. David Gavaghan	Co – Investigator	david.gavaghan@comlab.ox.ac.uk Tel: 01865 610667
Prof. Peter Kohl	Co – Investigator	peter.kohl@dpag.ox.ac.uk Tel: 01865 272114
Ben O'Steen	Lead Metadata Management	benjamin.OSteen@sers.ox.ac.uk Tel: 01865 283864
Kathryn Dally	Lead Policy Development	kathryn.dally@admin.ox.ac.uk Tel: 01865 280319
Luis Martinez-Uribe	Project Manager and Analyst	luis.martinez-uribe@oerc.ox.ac.uk Tel: 01865610607
	Systems Developer	
	IBM representative (by invitation)	
	Planning and Resource Allocation Section representative	

The Project Manager, reporting to the Head of Infrastructure Systems and Services (OUCS), will have the responsibility of day to day coordination of the project; including maintaining the project plan, and liaising between collaborators. The Analyst will lead the audit and requirements analysis workpackage and contribute to the economics and sustainability workpackage. Both the Library Services and the Research Services Office will manage analyst activities relating to metadata and policy framework respectively. A Systems Developer will conduct the technical analysis and lead the development of the technical implementation workpackages.

13. Programme Support

The project team requests support from the Programme Manager in the following areas:

- Identifying external evaluators for the formative evaluation that will take place half-way through the project.
- Alerting and introducing the project to other relevant activities.
- Helping to define the aims and objectives of the two workshops as well as promoting them.

14. Budget

See Appendix A.

Detailed Project Planning

15. Workpackages

See Appendix B.

16. Evaluation Plan

The project has set aside funds for a formative evaluation around half-way through the project by an external evaluator. This will help ensure that the deliverables serve the purpose for which they are intended; that the project has developed an appropriate framework that integrates the requirements of the research groups; the emerging institutional policy for data management and sustainability model; and the deployment of tools to serve the collated requirements. In addition to the formative evaluation,

the project team will reflect on progress throughout the project and document this in the progress reports and final report.

Timing	Factor to Evaluate	Questions to Address	Method(s)	Measure of Success
Half-way (around end of 2009)	Aims and objectives	Does the project progress appropriately towards meeting its aims and objectives as expected at this time?	Review documentation	Meeting milestones and comparing progress to completed workpackage tasks
	Data audits	Has the data audits been completed successfully?	Interview Analyst and review documentation	The two research groups being audited and results recorded in report
	Technical requirements and workflow module development	To what extent has the first iteration of technical requirements and workflow module development taken place?	Interview systems developer and review documentation	The research groups' requirements have been turned into a first round of functional requirements and there has been progress in the development of the workflow module.
	Federated data storage and access infrastructure report	To what extent there has been progress on the production of a report on the strategic provision of data storage and access infrastructure?	Interview IBM representative and review documentation	There has been appropriate progress in the evaluation of possible solutions to address the infrastructure requirements and in the production of the report.
	Metadata Standards	Have metadata standards been identified and agreed to be used to describe, administer and preserve the research data?	Interview Lead Metadata Management and review documentation	There are appropriate metadata standards to describe the research data and the Project Working Group has approved them
	Policy Framework	To what extent have policies been mapped across the institution and outside and there has been appropriate progress into drafting an over-arching research data management policy framework?	Interview Lead Policy Development and review documentation	Different policies have been identified across Oxford and outside and there has been progress towards drafting an over-arching policy for research data management
	Economics and sustainability	To what extent have different economic models for managing research data have been identified and are been applied/adapted to the specific research data in this project?	Interview Project Manager	Economic models have been assessed and work progresses according to plan.

	Dissemination	Has internal and external dissemination through the different channels of communication run according to the dissemination plan?	Review documentation and interview Project Manager	The project website, the blog and the bookmarks are up and running; a workshop has taken place; papers or posters have been submitted to conferences.
	Overall progress	Has the work so far been successful?	Interview Project Manager and focus group with Project Working Group	Outcomes and outputs match the expectations from all project stakeholders and overall agreement that the project is running smoothly

17. Quality Plan

Timing	Quality criteria	QA method(s)	Evidence of compliance	Quality responsibilities
Output	Project website, blog and bookmarks			
End of April 2009	Adherence to standards and specifications, well written content	Follows W3C standards and is reviewed by Project Working Group	Validation checks are passed and signed-off by Project Working Group	Project Manager, Project Working Group
Output	Project workshops			
October 2009 & June 2010	Well attended and positive feedback	Evaluation and feedback from attendees	Evaluation report and positive feedback from delegates	Project Manager, Project Working Group and Programme Manager
Output	Data audits and requirements analysis report			
End of July 2009	Fitness for purpose, enough information to support work in WP3 and WP4	Evaluation process and timely delivery of outputs	Evaluation report and signed off by Project Working Group and Programme Manager	Analyst, Project Manager and Project Working Group.
Output	Metadata standards			
End of October 2009	Appropriate level of detail to describe, administer and preserve as well as to allow reuse	Evaluation and user testing	Evaluation report, positive user feedback and signed off by Project Working Group	Lead Metadata Management, Analyst, Project Working Group
Output	Curatorial workflow module			
End of July 2010	Fitness for purpose, adherence to open standards	Evaluation and user testing	Evaluation report and positive user feedback	Systems Developer, Project Manager, Project Working Group

Output	Federated data storage and access infrastructure report			
End of July 2010	Fitness for purpose	Evaluation and is reviewed by Project Working Group.	Evaluation report and signed off by Project Working Group	IBM representative, Project Working Group
Output	Overarching research data management framework			
End of April 2010	Fitness for purpose	Evaluation and is reviewed by Project Working Group	Evaluation report and signed off by Project Working Group	Lead Policy Development, Analyst, Project Working Group
Output	Report on the application of costing frameworks			
End of August 2010	Fitness for purpose	Evaluation and is reviewed by Project Working Group	Evaluation report and signed off by Project Working Group and Programme Manager	Project Manager, Project Working Group
Output	Final report			
End of September 2010	Fitness for purpose	Reviewed by Project Working Group and Programme Manager	Signed off by Project Working Group and Programme Manager	Project Manager, Project Working Group

18. Dissemination Plan

Timing	Dissemination Activity	Audience	Purpose	Key Message
By month 3	Project website, blog and bookmarking site	Any interested parties	Raise awareness, promote and engage	Aim and objectives, plans and progress
By month 8 and month 18 respectively	Two workshops	Internal/external stakeholders and wider community	Engage	Institutional data management and curation is a cross-agency activity
As opportunities arise	Articles and other publications in relevant journals	External stakeholders and wider community	Inform and promote	Achievements and experiences
As opportunities arise	Conference presentations or posters submitted to relevant events	External stakeholders and wider community	Engage and promote	Achievements, experiences, outputs and outcomes
As opportunities arise	Internal dissemination	Internal stakeholders	Raise awareness, educate and engage	Appropriate mechanisms to manage and curate research data need to be in place in Oxford
Throughout the project	Reports and other documents	Any interested party	Inform	Outputs

Throughout the project	Communication with other related projects	Other JISC projects	Engage	Share experience and expertise
As necessary	Involvement in JISC programme events	Dependent on opportunities	Dependent on opportunities	Dependent on opportunities

19. Exit and Sustainability Plans

Project Outputs & Outcomes	Action for Take-up & Embedding	Action for Exit
All forms of reports including audit findings, filestore provision, costing frameworks, progress, evaluation and final reports	Deliver to appropriate individuals and deposit copies on the Oxford Research Archive (ORA)	Ongoing preservation and maintenance for ORA.
Curatorial workflow module	Ensure that the technologies developed meet research groups' needs and made the software available under an open source software license	Clear responsibilities for continued development and maintenance of the technologies and services developed
Project website, blog and bookmarks	Negotiate links to website from other key University web pages	Ensure editorial, hosting and maintenance responsibilities.
Overarching framework for the management research data generated in Oxford	Relevant internal stakeholders groups are involved in the production of the framework	Ensure responsibilities for the continued development of the framework seeking approval from relevant internal groups in Oxford
Curated research data generated by research groups	Research data curated with enough information about provenance, ownership and preservation related actions	Clear responsibilities for the continued curation of the research data

Project Outputs	Why Sustainable	Scenarios for Taking Forward	Issues to Address
Curatorial workflow module	The project expects that the approach used to preserve data can scale to address the data preservation challenges of other research groups in Oxford.	The technologies and services developed are useful to other research groups in the University of Oxford.	Further development, maintenance and support responsibilities.
Overarching policy framework for the management and preservation of research data generated in Oxford	The overarching framework represents a first stage towards developing and approving an institutional policy for the management and preservation of research data.	A consultation with relevant stakeholders in Oxford starts to approve the research data management and preservation framework.	Further development of framework and institutional approval.
IBM report on strategic provision	This report is a first stage towards	The report contains recommendations on how	Recommendations need to be taken

Project Acronym: EIDCSR
 Version: 2
 Contact: Luis Martinez-Urbe
 Date: 28 July 2009

of data storage and access infrastructure	addressing these requirements identified in previous projects and by other services.	to best address those two requirements	forward
Curated research data generated by research groups	The funding agency BBSRC requires the research data to remain accessible for ten years after the completion of the project.	The data with the appropriate metadata is archived in the HFS and the metadata is also on the Fedora DAMS.	Further reviewing and maintenance of the research data on a regular basis.

Appendix A. Project Budget

Directly Incurred Staff	Apr09– Mar10	Apr10 – Mar11	TOTAL £
Total Directly Incurred Staff (A)	£ 99,782.00	£ 37,079.00	£ 136,861.00
Directly Incurred Non-Staff	Apr09– Mar10	Apr10 – Mar11	TOTAL £
Total Directly Incurred Non-Staff (B)	£ 13,391.00	£ 9,936.00	£ 23,327.00
Directly Incurred Total (C) (A+B=C)	£ 113,173.00	£ 47,015.00	£ 160,188.00
Directly Allocated	Apr09– Mar10	Apr10 – Mar11	TOTAL £
Directly Allocated Total (D)	£ 33,432	£ 15,474	£ 48,906
Indirect Costs (E)	£ 104,072.00	£ 40,202.00	£ 144,274.00
Total Project Cost (C+D+E)	£ 250,677.00	£ 102,691.00	£ 353,368.00
Amount Requested from JISC	£ 159,364.00	£ 65,636.00	£ 225,000.00
Institutional Contributions	£ 91,261.00	£ 37,107.00	£ 128,368.00
Percentage Contributions over the life of the project	Partners 36%	JISC 64%	Total 100%
No. FTEs used to calculate indirect and estates charges, and staff included	No FTEs: 1.95	All DI and DA staff.	