

<b>Cover Sheet for Proposals</b> <i>(All sections must be completed)</i>		<b>JISC</b> Digital Repositories Programme	
<b>Project Area:</b> c) Pilot Implementations of the data audit framework			
<b>Name of Lead Institution:</b>		UCL	
<b>Name of Proposed Project:</b>		UCL DAF project	
<b>Name(s) of Project Partner(s):</b>			
<b>Full Contact Details for Primary Contact:</b>			
Name: Martin Moyle			
Position: Digital Curation Manager, UCL Library Services			
Email: m.moyle@ucl.ac.uk			
Address: Science Library, UCL, Gower Street, London WC1E 6BT			
Tel: 020 7679 4351			
Fax: 020 7679 7373			
<b>Length of Project:</b>		7 months	
<b>Project Start Date:</b>	1 April 2008	<b>Project End Date:</b>	31 October 2008
<b>Total Funding Requested from JISC:</b>		£ 30,000	
<b>Total Institutional Contributions:</b>		£ 37,268	
<b>Outline Project Description</b>			
UCL will test the Data Audit Framework (DAF) across a representative range of UCL academic departments, building on existing work, expertise and contacts within UCL. A full report on the exemplar audit will be made available to the HE community.			
<b>I have looked at the example FOI form at Appendix A and included an FOI form in the attached bid (Tick Box)</b>		YES	
<b>I have read the Circular and associated Terms and Conditions of Grant at Appendix B (Tick Box)</b>		YES	

## Response to JISC Circular 5/07: Digital Repositories

### UCL (University College London): Exemplar Implementation of the Data Audit Framework

#### 1. Introduction

This proposal is submitted by UCL Library Services in response to the JISC Circular 5/07: Call for Projects in Digital Repositories. The proposal is submitted to theme B (ii): Exemplar Implementations of the Data Audit Framework.

UCL will test the Data Audit Framework (DAF) across a representative range of UCL academic departments, building on existing work, expertise and contacts within UCL. A full report on the exemplar audit will be made available to the HE community.

The proposal is for a 7-month project at a total cost of £ 67,268. The total funding sought from the JISC is £ 30,000.

#### 2. Strategic Context

Primary research data represents the building blocks of scholarly research. Across all major disciplines – Medicine, Engineering, Physics, Social Sciences, the Arts and Humanities – primary data forms the basis for the secondary research outputs in those disciplines in the form of monographs, journal articles, Reports and Conference Papers.

The embedding of repositories in UK HEIs has brought greater focus onto secondary research outputs. Repositories give such publications greater visibility. With material such as research Ph.D theses, these materials receive far more consultations through being based in a repository than the paper copy does on a library shelf.

With some notable exceptions, repositories have to date not tackled the storage, preservation and sharing of primary research data. This is an area of academic activity that is ripe for study. It is not clear to academics who owns the primary research data they and their team collect. Nor is there a real awareness how such data should be curated and preserved in the long term. Who is responsible for doing this? Where should it be stored? How much will it cost? In principle, primary research data which is properly curated can be re-used by any researchers or research teams that have access to that curated material – thus saving money by not having to replicate data collection.

Few, if any, UK Universities know what primary research data they hold, where it is stored, who (if anyone) is caring for it, and what the best way of preserving that data in the long term is. UCL is a university now ranked 9<sup>th</sup> in the world league table of universities.<sup>1</sup> It has a strong research record in all major academic areas of study. The motivation for UCL Library Services to submit this bid to the JISC for partnership in developing a Data Audit Framework is the need for UCL to be clear what primary data it holds, how it is managed and made available for other researchers, and whose responsibility it is to undertake the long-term curation of this material.

The output of the work will provide UCL and the HE community with a collection of generic outputs and outcomes, which can be shared as models of good practice:

- a map of current collections of primary research data held in an institution
- a methodology for analysing and assessing the collections and formats of primary research data
- the roles and responsibilities of individuals connected with the data's
  - collection
  - storage

---

<sup>1</sup> See [http://www.timesonline.co.uk/tol/life\\_and\\_style/education/article2827773.ece](http://www.timesonline.co.uk/tol/life_and_style/education/article2827773.ece).

- interpretation
- the workflows which surround primary research data in terms of
  - curation
  - long-term preservation

UCL's participation as a partner in the development of a Data Audit Framework will bring benefit, not simply to UCL, but to all UK HEIs with a strong research record.

### **3. Commitment to Data Curation at UCL**

UCL has already begun to address the challenges of digital curation, including the active and sustainable management of its research data. By way of illustration, three existing commitments are outlined here.

#### **3.1. UCL Digital Curation Working Group**

UCL Information Strategy Committee has established a Working Group to take forward the digital curation agenda within UCL. The Group meets termly, is chaired by the Director of UCL Library Services, and reports to UCL's high-level Information Strategy Committee. The UCL DAF Proposal is entirely consonant with the aims and objectives of the UCL Working Group on Digital Curation. Specifically, the role of the Working Group is to:

- lead on digital curation issues for UCL
- identify best practice in the curation of scholarly and administrative materials in digital formats, including primary data
- disseminate best practice in digital curation across UCL
- map the growth of digital objects in UCL and prepare UCL to undertake their long-term digital curation
- sponsor and pilot new initiatives, technology and processes in digital curation within UCL

The membership of the UCL Working Group is drawn from UCL's central services and a range of academic departments. The Project Board for the UCL DAF project will be drawn from members of this Working Group (see 5.4).

#### **3.2. EMBRACE Project**

The EMBRACE<sup>2</sup> (Embedding Repositories and Consortial Enhancement) project is co-funded by the JISC and UCL, on behalf of the SHERPA-LEAP<sup>3</sup> Consortium. A substantial proportion of the project's resource is to be given over to an investigation of drivers for, and barriers to, the embedding of digital repositories in institutional strategy. The project will examine senior-level attitudes and practices in at least three SHERPA-LEAP HEIs of different size and mission, specifically concerning the institutional stewardship of research eprints, electronic theses, learning objects and courseware, and primary data. One of the project outputs will be a strategic "toolkit", which will identify the questions, issues and measures to be considered when a proposer is pursuing a strategic commitment to resourcing and sustaining repositories of digital assets within a HE institution. The work of EMBRACE in support of the business case for data curation at an institution will be complemented by UCL's participation in the development of an audit framework for institutional data.

#### **3.3. LIFE2 Project: Case Study on Primary Data**

UCL jointly leads, with the British Library, the LIFE<sup>4</sup> project, which has developed a methodology to model the digital lifecycle and calculate the costs of preserving digital information in the long term. The LIFE Team is running a number of case studies during the project's second phase (LIFE2), one of which is based on primary data owned by the Medical Research Council. This collection is a prospective, longitudinal birth cohort study, and is one of the longest running large-scale studies of human development in the world. UCL's participation in this investigation dovetails cleanly with this proposal for UCL to participate in the development of a Data Audit Framework.

---

<sup>2</sup> EMBRACE Project: <http://www.sherpa-leap.ac.uk/embrace>

<sup>3</sup> SHERPA-LEAP: <http://www.sherpa-leap.ac.uk>

<sup>4</sup> LIFE/LIFE2: <http://www.life.ac.uk>

#### 4. UCL: Other Relevant Experience

UCL Library Services is able to demonstrate a successful track record in leading and contributing to JISC projects relevant to the UCL DAF proposal. In particular:

##### 4.1. Leadership of JISC projects

UCL Library Services recently led, with the British Library, the first phase of the LIFE project (see 3.3), and led an evaluation study leading to a report on options for an Electronic Thesis Service.<sup>5</sup> UCL is currently leading the RIOJA (Repository Interface to Overlaid Journal Archives)<sup>6</sup> project, the EMBRACE project (see 3.2), and co-leading the second phase of the LIFE project, LIFE2 (see 3.3).

##### 4.2. Contribution of case studies to JISC projects

UCL Library Services has recently contributed case studies to two JISC projects. The ShibboLEAP<sup>7</sup> project investigated aspects of federated access management using Shibboleth technology. UCL Library Services, working with UCL Information Systems, contributed a case study of Shibboleth IdP implementation at UCL. The Identities Project<sup>8</sup> aimed to investigate and document the Identity Management situation in UKHE: UCL Library Services undertook an in-depth audit of Identity Management practices at UCL and contributed a full report on the findings.

#### 5. Overall Approach

##### 5.1. Project Methodology

It is noted that the UCL DAF project will be part of a suite of national work, which has a significant level of central coordination. It is therefore anticipated that the exact project methodology will be determined in the early stages of the project itself. Within that constraint, UCL is able to propose the following:

###### 5.1.1. Work Packages

The project will run for 7 months, and it is anticipated that the work will fall into 5 main areas:

- Project Management (see 5.3).  
Timescale: months 1-7.
- Preparatory work - liaison with the DAF project and other participating institutions; initial development of DAF; agreement of scope, timescales and deliverables.  
Timescale: months 1-2.
- Audit - internal planning and execution of UCL Data Audit.  
Timescale: months 3-5.
- Report - preparation of full report on UCL Data Audit.  
Timescale: months 6-7.
- Dissemination (see 8).  
Timescale: months 1-7.

At the outset of the project, a detailed Project Plan, including all appropriate Work Packages, will be drawn up, in conjunction with the plans of the DAF project and the other exemplar institutions.

###### 5.2.2. Audit participants

UCL has 72 academic departments staffed by 4,000 academic and research staff. UCL research spans all major academic disciplines. The scale and range of UCL research makes the University a good testbed for the Data Audit Framework; but it is doubtful that a comprehensive audit will be achievable within the time constraints of this project. It is therefore proposed that UCL should identify and work with a representative selection of academic departments for the exemplar audit. The selection will include departments with research interests in biomedicine, engineering, pure science, arts and humanities, social sciences, and environmental studies, and will include Departments with Research Groups

---

<sup>5</sup> Evaluation of Options for a UK Electronic Thesis Service: <http://eprints.ucl.ac.uk/archive/00002197/>

<sup>6</sup> RIOJA: <http://www.ucl.ac.uk/ls/rioja>

<sup>7</sup> ShibboLEAP: <http://www.angel.ac.uk/ShibboLEAP/>

<sup>8</sup> Identities Project: <http://www.angel.ac.uk/identity-project/index.html>

(such as High Energy Physics) whose data is currently stored away from UCL, as well as Departments with data held on institutional systems. Interdisciplinary research will be exemplified in the project by UCL Research Computing. It is expected that the precise number and research interests of departmental participants in the exemplar audit will be determined after further discussion with the JISC and the DAF project.

### 5.3.3. Timescales

To allow for a prompt start date, and to support flexibility in the timing of the project, UCL will aim to second an existing member of staff to this post. It is anticipated that further refinement of the project timescales will take place in discussion with the DAF project and the other exemplar projects.

For budgetary purposes, a start date of 01 April 2008 has been assumed (see 11).

## 5.2. Staffing

UCL will appoint a Project Officer to oversee the implementation of the DAF at UCL. The Project Officer will liaise on a day-to-day basis with the DAF project. He or she will also liaise directly with the UCL departments selected to participate in UCL's exemplar Audit, as well as with the other DAF exemplar institutions, and will be responsible for documenting UCL's implementation of the DAF.

## 5.3. Project Management

UCL will assign a Project Manager to ensure that the project is managed effectively and that all the project outputs are delivered to time and within budget.

A project management Work Package will be drawn up. Deliverables from this work package will include the detailed Project Plan (pending discussion with the main DAF team and other participants), meetings of the Project Board, and the coordination of project evaluation, dissemination and reporting activity. The Project Manager will be accountable to the Project Board for the progress of the project.

## 5.4. Project Governance

A Project Board will be convened to oversee the development of the project. It will meet at least twice in the course of the project. The role of the Project Board will be to monitor, review, assess and sign off the outputs of the project, to sponsor all dissemination activity, and to help to shape the work of the Project Officer, particularly by identifying appropriate academic departments for participation in the audit and ensuring that work progresses smoothly. Periodic reports will be made to UCL Information Strategy Committee, which meets fortnightly.

The Board will consist of the following members of UCL's Digital Curation Working Group:

- Dr Paul Ayris, Director of UCL Library Services and UCL Copyright Officer (Chair)
- Professor Jonathan Butterworth, High Energy Physics, UCL Physics and Astronomy
- Mr Robert Clark, Director of UCL Information Systems
- Professor Tim Cole, Paediatric Epidemiology and Biostatistics, UCL Institute of Child Health
- Mrs Jan Cropper, Group Manager IT Services, UCL Library Services
- Dr Clare Gryce, UCL Research Computing Coordinating Manager
- Mr Martin Moyle, Digital Curation Manager, UCL Library Services
- Dr Melissa Terras, UCL School of Library, Archive and Information Studies
- Dr Jeremy Yates, Astrophysics, UCL Physics and Astronomy

The Board will also have the following external member:

- Ms Helen Hockxx, PLANETS Project Manager, British Library

## 6. Risks

Risk	Probability	Impact	How managed
Inability to second a Project Officer	Low	Med	External recruitment.
Key staff leave during project lifetime	Low	Med	Secondment reduces this risk. If key staff leave early in project, re-recruitment will take place. If disruption is towards the end of the project, the UCL Digital Curation Team will complete the work.
Many UCL academic departments reluctant to participate	Low	High	Known enthusiasm for data curation in some departments makes this unlikely. Sponsorship of UCL Information Strategy Committee will ensure good levels of participation.
Cooperation of a UCL department is withdrawn	Low	Med	Department with similar academic interests will be found, with backing of UCL Information Strategy Committee.
Delays to development of Digital Audit Framework	Med	Med	UCL flexible over start date. Staffing of project by secondment may permit additional flexibility.

## 7. IPR

All published material produced by the project will follow current JISC guidelines for accessibility. Copyright of the appropriate author(s) and/or partner institution(s) will be asserted, but all such material will be covered by the appropriate Creative Commons licence allowing free non-commercial reuse.

## 8. Dissemination

The project team will be committed to disseminating the work of the project and its outputs, and will seek opportunities to present the work of the project at relevant national and international conferences, in conjunction with the DAF development team and the other DAF Exemplar Implementation projects. A Web site will be created to share information about the UCL DAF project. It will provide links to project reports and to copies of all publications and presentations deriving from the project. Additionally, UCL's final worked example and supporting documentation will be made openly and freely available through both UCL's Institutional Repository and the JISC's document repository. The Project Officer will be responsible for dissemination about the project within the UCL community. The Project Board will assist with the identification of dissemination opportunities and will sponsor all project dissemination activity.

## 9. Sustainability

The UCL DAF implementation is a timed project with a fixed set of deliverables, and the sustainability of its outputs does not, therefore, require further resources. The outputs will contribute to the future sustainability in general of data curation activities in the wider JISC community.

## 10. Summary of key outputs

The UCL DAF project will contribute the following key outputs for the long-term benefit of the JISC community:

- Participation in the iterative development of a Data Audit Framework
- A map of current collections of primary research data held across a wide range of research-active departments and interdisciplinary centres within an institution
- Identification of the roles and responsibilities of individuals connected with the data's collection, storage and interpretation



## Appendix A

### FOI Withheld Information Form

We would like JISC to consider withholding the following sections or paragraphs from disclosure, should the contents of this proposal be requested under the Freedom of Information Act, or if we are successful in our bid for funding and our project proposal is made available on JISC's website.

We acknowledge that the FOI Withheld Information Form is of indicative value only and that JISC may nevertheless be obliged to disclose this information in accordance with the requirements of the Act. We acknowledge that the final decision on disclosure rests with JISC.

<b>Section / Paragraph No.</b>	<b>Relevant exemption from disclosure under FOI</b>	<b>Justification</b>



Nike Holmes  
JISC Executive  
Northavon House  
Coldharbour Lane  
Bristol BS16 1QD

11 January 2008

Dear Nike,

**JISC Circular 5/07: Digital Repositories - Exemplar Implementations of the Data Audit Framework**

I am very pleased to submit the enclosed response outlining UCL's proposal to carry out an exemplar implementation of the Data Audit Framework.

Participation in the Data Audit Framework offers to enrich UCL's strong strategic commitment to the digital curation agenda. Research is increasingly driven by primary data, and the Data Audit Framework offers UK Universities an important mechanism for achieving clarity over what data they hold, and who is responsible for its long-term management and curation. The learning outcomes of the UCL exemplar implementation will be of considerable benefit to the HE community at large.

I confirm that UCL will meet all the responsibilities and resource commitments outlined in the proposal.

Yours sincerely,

A handwritten signature in black ink that reads 'Paul Ayris'. The signature is written in a cursive style and is positioned above a horizontal line that extends to the right.

Paul Ayris, MA, PhD