JISC Final Report

Project

Project Information			
Project Acronym	UCL DAF		
Project Title	UCL Data Audit Framework pilot implementation		
Start Date	18 August 2008	**End Date**	April 2009
Lead Institution	UCL Library Services		
Project Director	Paul Ayris		
Project Manager & contact details	Martin Moyle m.moyle@ucl.ac.uk / 020 7679 4351		
Partner Institutions	By terms of implementing the Data Audit Framework Methodology and collaborating with DAF: ▪ HATII University of Glasgow (http://www.hatii.arts.gla.ac.uk/index.html) and ▪ Digital Curation Centre (DCC) (http://www.dcc.ac.uk)		
Project Web URL	http://www.ucl.ac.uk/ls/data-audit		
Programme Name (and number)	Digital Repositories programme 2007-8		
Programme Manager	Dr Neil Jacobs		

Document

Document Name			
Document Title	*Final Report*		
Reporting Period			
Author(s) & project role	Panayiota Polydoratou, Project Officer & Researcher		
Date		**Filename**	
URL	http://eprints.ucl.ac.uk/15053		
Access			☐ General dissemination

Document History		
Version	**Date**	**Comments**
0a	March 2009	Draft to Programme Manager

UCL DAF pilot (UCL Data Audit Framework pilot implementation project)

Final Report

April 2009

Table of Contents

Acknowledgements

The UCL DAF pilot was funded by the JISC (Joint Information Systems Committee, UK). We would like to extend our thanks to all academic and research staff who participated in our pilot and contributed with useful information and suggestions.

We would also like to thank the following people for their support:

Professor Stephen Shennan (Director of the Institute of Archaeology)
Dr Katie Meheux (Librarian, Head Issue Desk, Institute of Archaeology)
Dr Mary Hilson (Head of Department (to December 2008), Department of Scandinavian Studies)
Dr Claire Thompson (Head of Department (since January 2009), Department of Scandinavian Studies)
Professor Jonathan Tennyson (Head of Department, Department of Physics and Astronomy)
Dr Jeremy Yates (Deputy Manager Research Computing and Honorary Senior Research Fellow, Department of Physics and Astronomy)
Dr Suzanne Beeke (Head of Department, Research Department of Language and Communication)
Dr Merle Mahon (Senior Lecturer, Research Department of Developmental Science)
Professor Ann Blandford (Director of UCL Interaction Centre)
Dr Andrew Faulkner (Head of Department, Research Department of Speech, Hearing and Phonetic Sciences)

Executive Summary

The UCL DAF pilot implementation project was funded by JISC for a period of seven months (August 2008-March 2009) to trial the implementation of the DAF methodology at UCL. The pilot was trialled across a range of departments and an interdisciplinary research centre at UCL. Specifically, the departments and centre that were approached are:

1. Department of Scandinavian Studies (Arts & Humanities)
2. Institute of Archaeology (Social & Historical Sciences)
3. UCL Interaction Centre (Interdepartmental and cross-faculty research)
4. Department of Language and Communication (Life & Medical Sciences)
5. Department of Speech, Hearing and Phonetic Sciences (Life & Medical Sciences)
6. Department of Physics and Astronomy (Mathematical & Physical Sciences)

This pilot project aimed at auditing primary research data only. For the purposes of the project, a definition of the characteristics of primary research data across disciplines and practices is as follows:

> Primary research data are data produced within the timeframe of a project/research work/lifetime. They are unprocessed (often referred to as raw data), original, generated by machines or humans and are regarded as the core of any research activity.

Overall, the 4-stage methodology proposed by the DAFD was followed. However, some allowances had to be made for the short timeframe and pilot nature of the project. Specifically, some selectivity in methodology had to be exercised in order to ensure useful outcomes for both the project partners and the UCL academic community.

Timing, clear scope and an institutional requirement for an audit were highlighted by the participants as crucial factors for success. Picking the right time for the audit (e.g. the Summer academic recess) could ensure staff engagement. Clear scope and linking the findings to a wider information strategy that might promise tangible results (e.g. a primary research data repository) was also highlighted as possible motivational factor for participation. An institutional requirement for an audit demonstrates, in some respect, a potential commitment to the implementation of the results.

Part of the exercise included conducting an online questionnaire survey and interviews with members of staff. Specifically 192 people were contacted using the e-informs software (received response by 57 people, ~30% response rate) and conducted interviews with 30 members of staff. Some information about the research profile of the participating departments based on publication outputs can be found in Appendix B. The project delivered: a website (http://www.ucl.ac.uk/ls/data-audit), a brief report to the project's partners describing the UCL DAF approach, a news item about the project that has been published at D-Lib Magazine (http://www.dlib.org/dlib/january09/01contents.html#IN-BRIEF, doi:10.1045/january2009-inbrief), a paper discussing some of the survey results, which has been accepted for publication at the ELPUB 2009 conference, and results from the questionnaire survey in the form of tabular data which are to be handed to the audited departments. A more detailed analysis of the interviews could produce further information about current data management and storage practices and requirements, back-up and archiving routines as well as potential uses of primary research data as indicated by the interviewees. Some of the issues that could be explored further are: examples of sharing of research data, assigning "value" indicators to data, practical issues (metadata and uploading of data, technical infrastructure)

The online tool was unavailable during the planning phase of the project and the audits were planned and executed with the .xls based audit forms. Therefore we could not test and provide feedback on the usability of the tool and how it would have been received by the academic and research staff. Overall, the DAF methodology was flexible, the coverage of support documents to consult for auditing research data was good, it created a feeling of bonding with the audited departments, and depending on the tool's interface and functionality, could potentially capture rich data (via the forms 3A&B). Limitations are discussed in the Implementation section.

Background

University College London is a global research university which encompasses 200 years of discovery and achievements and aims to transform the world. Research has been a core activity at UCL since the university's foundation in 1826 and is at the heart of its future strategy.[1] UCL counts amongst its strengths more than 4000 academic and research staff of which, about 17% have reached the top of their careers (i.e. professorship for academic staff) and are involved in the forefront of research.[2]

During 2006/2007, UCL staff have successfully secured research funds and excelled in scholarship. Income generated by research funds and grants accounted for one third of the overall UCL income. Recognition for the advancement of scholarship includes fellowships in Royal Societies and Academies, professional associations as well as 20 Nobel Laureates up to now that have been awarded to UCL staff since 1901, most recently to Professor Sir Martin Evans in the field of Physiology of Medicine, in 2007. Furthermore, recognition at an international level is demonstrated in the Times Higher Education (THE) rankings that list UCL 9th of the world top 200 universities.[3] Outputs from a recent bibliometric study placed UCL in the 2nd place of most productive research institutions in Europe, 3rd place as the most cited university and the UK University most cited by health researchers (van Raan, 2008).

The diversity, volume and breadth of the data produced by UCL scientists are further enhanced by the merits of interdisciplinary research which, UCL strongly promotes. It often involves collaboration with other universities and at international level and is made evident by the number and cross boundaries of the centres and institutes at UCL. Further information can be found at: http://www.ucl.ac.uk/research/departments/themes

UCL Library Services is in the core of the network that supports the university's vision. The Library states its commitment to support excellence of teaching, learning, research and clinical practice by providing a high quality, integrated and innovative service.[4]Furthermore, and in compliance with advances in the provision of access to information, UCL Library Services is committed to the responsible, long-term stewardship of all its digital assets, whether born-digital or created through digitisation. The Library works with the Information Strategy Committee and the Information Systems and Media Resources to align the development of portals with existing tools to realise our mission. In further support, an established Working Group on Digital Curation:

- leads on digital curation issues for UCL
- identifies best practice in the curation of scholarly and administrative materials in digital formats, including primary data
- disseminates best practice in digital curation across UCL
- maps the growth of digital objects in UCL and prepare UCL to undertake their long-term digital curation
- sponsors and pilots new initiatives, technology and processes in digital curation within UCL

Participation in the Data Audit Framework Development project and implementation of the methodology falls exactly in line with UCL's mission and commitment to support research and teaching and the recommendations made by Liz Lyon (2007) in the JISC commissioned report "Dealing with Data". UCL aimed to gain an understanding of: a) what primary research data are held at designated faculties and departments, b) how the data are managed, and c) where lies the responsibility for their long term curation.

The findings from the UCL DAF implementation may be beneficial to other pilots undertaking the DAF methodology, the UCL academic community, funding organisations and other Higher Education Institutions. Furthermore, the auditing process may prove whether there is ground to explore the new roles for scientists and librarians in data care and handling as discussed by Swan and Brown (2008).

[1] UCL vision and strategy. Information available at: http://www.ucl.ac.uk/research/vision/
[2] UCL facts and figures. Information available at: http://www.ucl.ac.uk/research/facts/
[3] Times Higher Education (2007). World University Rankins: top 200 world universities. Information available at: http://www.timeshighereducation.co.uk/hybrid.asp?typeCode=144
[4] UCL Library Services e-strategy. Information available at: http://www.ucl.ac.uk/Library/e-strategy.shtml

The outputs of research activity are published records, most commonly in the form of journal articles, books, reports, etc. However, the vast majority of the data that are produced during the research process never reach publication stage. There is an increasing recognition of the potential benefits of such data both to the wider research community and to society in general.

Developing systems and services for the effective and efficient management of research data as well as addressing issues around their long term curation is an area of increasing activity in UK Higher Education. Recent examples of such activity include projects that currently run under the Joint Information Systems Committee (JISC) Digital Repositories programme 2007-9 such as the Data Audit Framework Development and a further study for identifying the benefits of curating and sharing research data, and the joint RLUK-RUGIT UK Research Data Service Feasibility study. Moreover, studies of the complex relations and issues around the generation, management, curation, use and value of research data (Swan & Sheridan, 2008; Lyon, 2007; Swan; NSB, 2005; MacLeod & Childs, 2003) have already started to pave the way for further research in this area.

The UCL Data Audit Framework pilot implementation project (UCL DAF) is part of the activity in this area.

Aims and Objectives

The UCL Data Audit Framework pilot implementation project aimed to implement the Data Audit Framework methodology which has been developed for auditing research assets in Higher Education Institutions. The aims of the UCL DAF project were to contribute to the iterative development of the Data Audit Framework and to collate information about the research assets generated, held and managed by academic and research staff at UCL. In particular, the objectives are:

a) explore the implementation of the methodology developed by the Data Audit Framework Development project (DAFD) for auditing research assets at Higher Education Institutions,
b) document, discuss and report issues and lessons learned from the pilot implementation with the members of DAFD and the other pilot implementation projects,
c) collate information about research assets and data management practices at designated departments/centres/institutes at UCL and
d) share the findings with the academic community and beyond.

Methodology

Approach on the implementation of the DAF project methodology

Overall, we aimed to follow the 4-stage methodology proposed by the DAFD, with allowances to be made for the short timeframe and pilot nature of the project - some selectivity in methodology had to be exercised in order to ensure useful outcomes for both the project partners and the UCL academic community.

Target population

The audits were planned with the target population of academic and research staff in mind, including those holding honorary and affiliated agreements with UCL. We contacted academic and research staff at 5 departments and an interdisciplinary research centre at UCL. Specifically, the departments and centre that were approached are:

1. Department of Scandinavian Studies (Arts & Humanities)
2. Institute of Archaeology (Social & Historical Sciences)
3. UCL Interaction Centre (Interdepartmental and cross-faculty research)
4. Department of Language and Communication (Life & Medical Sciences)
5. Department of Speech, Hearing and Phonetic Sciences (Life & Medical Sciences)
6. Department of Physics and Astronomy (Mathematical & Physical Sciences)

Some information about the research profile of the participating departments based on publication outputs has been grouped together and presented in a table available in Appendix B.

Timing and scope
The launch of the project coincided with the beginning of the academic term at UCL. This is a very busy time for both academic and research staff. Initial meetings with key contacts at the targeted departments suggested moving the audits further into the term to ensure that some members of staff would consider participating in the audits.

Taking into account the short duration of the project and limitations imposed by the timing of the project's launch, it was decided to introduce a questionnaire survey at this point (Appendix A) and plan the interviews (from which we hoped to obtain more detailed information about the research data) later in the term. The questionnaire survey was structured around the Form 2 (inventory of assets) of the methodology. The survey aimed to serve the following purposes: raise awareness about the project at the designated departments, gather information about types of research assets within discipline, facilitate quick completion of Form 2 of the DAF methodology, and act as means to plan interviews.

A hundred and ninety two (192) people[5] were invited to participate in the questionnaire survey which was administered using the e-informs software. The survey received responses from 57 people, approximately ~30% of those contacted and collected 32 examples of primary research data. Thirty (30) people were interviewed (11 volunteered via the survey).

Given the breadth and volume of research data generated by UCL, this pilot project aimed to audit primary research data only. For the purposes of the project, a definition of the characteristics of primary research data across disciplines and practices was as follows:

> Primary research data are data produced within the timeframe of a project/research work/lifetime. They are unprocessed (often referred to as raw data), original, generated by machines or humans and are regarded as the core of any research activity.

Implementation

In this section we report on lessons learned from the implementation of the DAF methodology.

Overall, the methodology was flexible. There were no restrictions in moving back and forth in stages, filling in as much information in the forms as becomes available. The coverage of support documents to consult for auditing research data was good and it created the feeling of bonding with the audited department due to the number and variety of people - holding different roles in the management of research data - that were allowed to be involved. However, the author felt that there were some limitations when it came to the scope, classification of research assets and the forms' length and terminology. Furthermore, the unavailability of the online tool made it difficult to test the final stages of the methodology adequately and provide feedback about its usability. We report our experience below:

Planning the audit
Planning the audits required a fair amount of time to be spent collecting information about the research conducted at the audited departments (e.g. number of staff by dept/web page, number of research projects/income generated, outputs of research/presence on Eprints, etc.) and arrange meetings with champions in the departments. The resources that were consulted were: UCL Research Publications, UCL Eprints, annual reports, departmental web pages (annual reports, history, etc.) and individual staff web pages. Also, planning and meeting with key contacts at the departments and getting approval for trialling the data audit required time, definition of scope and preparation for

[5] The Department of Scandinavian Studies decided to opt out from the questionnaire survey and participate in interviews only. Therefore, the number of staff at the department is not included in the questionnaire survey sample mentioned here.

the methods that were going to be used to collect the data (e.g. survey, interviews). The author felt that the methodology lacked guidance as to how to define the audit's scope and also how to practically identify research data workflows. This would have been very helpful given the complex nature of primary research data and their dependency on associated systems, services and networks. For example, Buchanan and Gibb (2007) write that from "*an information management perspective, key data*[6] *concerns are typically associated with data protection/storage, and records management and regulatory compliance*". Previous work in this area by MacLeod and Childs (2003) reports that research records comprise records of (i) the research process, (ii) the outcomes or products, (iii) the management of the process and (iv) the primary and analysed research data. The author understands this as a reference to wider information architecture comprising both records and the systems that generate, manipulate, manage and preserve them.

Although the methodology is flexible in the approach of the audit (top-down, across, hierarchical) it does not give guidance as to how to define the scope opposed to other information audit methodologies such as those discussed in Buchanan and Gibb (2007, 2008a, 2008b), e.g. through the provision of a scope matrix. Furthermore, the author was unable to meet the estimated times listed in the methodology. Specifically, getting hold of very busy academic and research staff at the beginning of the academic year proved difficult to do in the prescribed estimates of time (2-4 days spread over a period of 3 weeks). It was suggested by several members of staff from the audited departments that the best time for an audit would have been the summer academic recess.

Finally, a substantial amount of time went into the preparation and testing of the online questionnaire survey and the interviews.

Business case and expected outcomes

Defining the scope of the audit so that it appeals to members of staff in different research fields and engage in it was the most difficult task in the implementation of the methodology. It proved difficult to build a business case for the audit based on the methodology alone (see some of the comments from members of staff in Appendix D) and without being able to link the audit to a wider information strategy or policy requirement. The term "Research Assets" means different things to scholars/scientists in different fields and there seems to be an underlined assumption in the methodology that there is often an agreed, standardised management practice applied to research data prior to the audit. This is not true and in practice, research data management appeared to vary across departments and be project- and researcher-specific. As 5 audits were attempted in parallel it was not possible to estimate the exact time spent on each of them. For example, one department opted out from the questionnaire survey all together; all departments had specific timeframe requests for the interviews (e.g. taking place during the reading week, etc.).

Forms and online tool

During the planning of the audits the online tool was not available for use and therefore the audits were planned and executed using the paper version of the forms (for demonstration during interviews) and the electronic version (.xls file for completion). This fact restricted the audits as a) we were unable to test the tool's usability and acceptance by the audited departments and b) we were unable to demonstrate quick data collection and compilation of an inventory of assets for promotional purposes. Furthermore, taking into consideration that a fair amount of information needed to be provided by the members of staff themselves (in particular for part of the form 3A and 3B), the length, terminology and applicability of the forms were not regarded positively by the academic staff. The methodology assumes that there is a level of data management practiced already. But that is not always the case particularly, with types of research data that come in various formats (images, text, maps, databases, etc.) and are not always digitised. In some research fields the forms were deemed inappropriate due the nature of the research data (see comments in Appendix D)

This has been reflected in the data collection process. The quick filling of form 2 via the questionnaire survey provided an overview of types of primary research data. However attempts to encourage the completion of form 3A/B proved fruitless (only 5 of the 30 people who were interviewed returned a completed form giving examples of 1 or 2 of their data sets/files only).

[6] [as opposed to information]

Identifying and classifying data assets
Primary research data come in various formats and can be quantitative or qualitative. However, they do not have intrinsic value until they can be exploited. The range of values that primary research data can have was difficult to classify using the values 'minor', 'important' and 'vital' attributes suggested in the methodology. That was for various reasons: it was not always clear who owned the data (funder, researcher, university, etc.), the data may assume different values during the lifecycle of the project and after the project, the data may assume different values when they are stripped from the context that they were generated, etc.

Assessing the management of data assets and making recommendations
Members of staff at all audited departments were genuinely concerned about the management of their research data and felt they had a moral responsibility for the storage, preservation and access (current and future) to the data. Scientists in some fields actively share their data. Re-use of their own data is something the academic/research staff already practise and were particularly engaged when it came to potential use of their data by scholars in other fields (e.g., use of archaeological images by staff researching Scandinavian studies, use of videos that record verbal communication of subjects suffering a particular condition by scholars involved in clinical research but in a different research area, use of numeric data for text mining, etc.). The ownership of data, copyright and access to the data are areas that are still unclear and staff would welcome advice and support. Issues of data and access management for research assets belonging to retiring members of staff and/or also for members of staff who have left the University or are deceased were also raised.

Outputs and Results

The project aimed at completing a pilot implementation of the DAF methodology and to contribute to the development of the framework by providing feedback and sharing findings from its implementation. The following outputs were produced:

1. Project reports (project plan, interim and final report)
2. Progress report that listed the UCL DAF approach and some initial thoughts about strong/weak aspects of the methodology. That brief report was shared with the pilot implementation projects.
3. Questionnaire survey results (generated in the form of tabular data or a brief report if time permits)
    a. The results from the questionnaire survey were produced in the form of tabular data and will be shared with the department champions.
4. Paper(s) for submission to peer reviewed journal and/or international workshops/conferences
    a. A paper discussing the approach of the UCL DAF project on the trialling has been accepted at the ELPUB 2009 conference (http://www.elpub.net)
    b. A paper discussing the results from the questionnaire survey and interviews with the members of UCL staff is in preparation for submission to a peer reviewed journal. A more detailed analysis of the interviews would produce further results on data types, storage management and requirements, back-up and archiving practices and describe potential uses of research data as indicated by the interviewed staff.
5. News articles
    a. A news item about the UCL DAF project has been published in the DLib magazine (Polydoratou, P. (2009). The *UCL Data Audit Framework (DAF)* pilot implementation project. D-Lib, January/February, Vol.15 (1). Available at http://dlib.org/dlib/january09/01contents.html#IN-BRIEF)
6. Project Web site (http://www.ucl.ac.uk/ls/data-audit)

Outcomes

The aims of the UCL DAF project were:

a) explore the implementation of the methodology developed by the Data Audit Framework Development project (DAFD) for auditing research assets at Higher Education Institutions,

b) document, discuss and report issues and lessons learned from the pilot implementation with the members of DAFD and the other pilot implementation projects,
c) collate information about research assets and data management practices at designated departments/centres/institutes at UCL and
d) share the findings with the academic community and beyond.

Results from the questionnaire survey were produced in the form of tabular data and will be presented to the audited departments alongside with a brief description of the findings and recommendations. A brief overview of types of primary research data has been captured via the survey and Form 2 of the methodology. The interviews provided some useful information about research workflows by scientists in different fields and their concerns and suggestions regarding current data management practices. An introduction to our approach on the implementation of the DAF methodology has been published in D-Lib Magazine and a paper has been accepted at the ELPUB conference. A more detailed analysis of the interviews may provide complementary information about storage practices and requirements, back-up and archiving and potential uses of primary research data.

Conclusions

The UCL DAF pilot project trialled the implementation of the DAF methodology across 5 departments and an interdisciplinary centre at UCL. 192 people were contacted in total, via an online questionnaire survey. Responses were received from 57, and 30 people were interviewed (11 identified via the survey, 19 via personal correspondence). The questionnaire survey produced an inventory of 23 examples of research data and 8 detailed examples of completed forms 3 (by 5 people).

Overall, the author found the exercise useful and the interviews provided sufficient information about current data management practices and expectations. In brief, the author lists what she regarded as positive and weak factors of the DAF methodology:

Positive/Encouraging factors

- Flexibility. No restrictions in moving back and forth in stages, filling in as much information in the forms as becomes available.
- Coverage of support documents to consult for auditing research data.
- Created the feeling of bonding with the audited department due to the number and variety of people - holding different roles in the management of research data - that were allowed to be involved.
- Potential to capture rich data (via the forms 3A&B).

Weak/Restrictive factors

- Scope. The term “research assets” does not necessarily relate to research data only. Very general, unless defined for a purpose the staff did not feel as if they were contributing to a tangible result with potential benefits for the academic community and the public.
    - Lack of tools and techniques to assist in scope definition, e.g. a scope matrix.
- Classification. A potential value attribution in relation to benefits for the institution does not necessarily reflect the research process and value that academic and research staff assign to their data based on research process stage and individuals.
- Forms’ length and terminology appeared confusing and complicated.

Recommendations

Metadata/Forms

Explore means for automatic creation and/or completion of the forms (e.g. explore the feasibility of embedding the process of describing research assets into the research proposal submission; use of institutional repositories to parse funders‘ publicly available data; staff update once production of data has taken place, etc.). This would be regarded favourably by academic/research staff.

Classification/value
Explore the development of a model that allows classification of research data at various stages in their lifecycle (e.g. throughout the project, after the project, etc.)

Explore model with scope to delineate value to different parties (e.g. value for researcher, value for the public, value for policymaker, etc.)

Ensure that the audit is part of a wider information strategy and is repeated regularly.

Training in data management techniques for researchers.

References

References to projects and reports

Joint Information Systems Committee. Digital Repositories programme 2007-8. Information available at: http://www.jisc.ac.uk/whatwedo/programmes/digitalrepositories2007.aspx (Last accessed 18/02/2009)

Buchanan, S and Forbes Gibb (2007). The information audit: role and scope. International Journal of Information Management, 27, pp. 159-172. doi:10.1016/j.ijinfomgt.2007.01.002

Buchanan, S and Forbes Gibb (2008a). Methodology selection. International Journal of Information Management, 28, pp. 3-11. doi:10.1016/j.ijinfomgt.2007.10.002

Buchanan, S and Forbes Gibb (2008b). The information audit: theory versus practice. International Journal of Information Management, 28, pp. 150-160. doi:10.1016/j.ijinfomgt.2007.09.003

Data Audit Framework Development (DAFD). Information available at: http://www.jisc.ac.uk/whatwedo/programmes/digitalrepositories2007/dataauditframework.aspx (Last accessed 18/02/2009)

RLUK-RUGIT UK Research Data Service Feasibility Study. Information available at: http://www.ukrds.ac.uk/ (Last accessed 18/02/2009)

Fry, J; et al. (2008). Identifying the benefits of curating and sharing research data. Available at: http://jisc.ac.uk/whatwedo/programmes/digitalrepositories2007/databenefits.aspx (Last accessed 18/02/2009)

Swan, A. and B., Sheridan (2008). The skills, role and career structure of data scientists and curators: an assessment of current practice and future needs, Key Perspectives Ltd. Available at: http://ie-repository.jisc.ac.uk/245/ (Last accessed 18/02/2009)

Lyon, L. (2007). Dealing with data: roles, rights, responsibilities and relationships. UKOLN. Available at: http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing_with_data_report-final.pdf (Last accessed 18/02/2009)

National Science Board. Long-lived digital data collections enabling research and education in the 21st century. Technical Report NSB-05-40, National Science Foundation, September 2005. Available from: http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf (Last accessed 18/02/2009)

MacLeod, J. and Childs, S. (2003). Managing primary research data and records in HE institutes, Northumbria University; Information Management Research Institute. Available at: http://www.northumbria.ac.uk/static/5007/ceispdf/resrecrep.pdf (Last accessed 18/02/2009)

Data Audit Framework Methodology. Information available at: http://www.data-audit.eu/methodology.html (Last accessed 18/02/2009)

Humanities Advanced Technology and Information Institute (HATII). Information available at: http://www.hatii.arts.gla.ac.uk/ (Last accessed 18/02/2009)

Digital Curation Centre (DCC). Information available at: http://www.dcc.ac.uk/ (Last accessed 18/02/2009)

van Raan, A.F.J. (2008). Bibliometric statistical properties of the 100 largest European research universities: Prevalent scaling rules in the science system. Journal of the American Society for Information Science and Technology, **59**(3): p. 461-475. Pre print available at: http://arxiv.org/abs/0704.0889 (Last accessed on 10/03/2009)

Appendix A: Questionnaire

ABOUT YOU

Q1. Please let us know your Department:
WRITE IN

Q2. If applicable, please let us know your research group:
WRITE IN

Q3. From the following options, please select the one that best describes your current role:
TICK ONE ONLY

- ○ 1 Academic Staff (Professor, Reader, Senior Lecturer, Lecturer, etc.)
- ○ 2 Research Staff (Senior Research Fellow, Research Fellow, Research Associate, Research Assistant, etc.)
- ○ 3 Visiting/Honorary staff
- ○ 4 Other (please specify)

Q4. Are you currently actively involved in any research work?
TICK ONE ONLY

- ○ 1 Yes
- ○ 2 No, my time is dedicated to other academic/educational activities at the moment
- ○ 3 Research is not part of my current role (please provide any specification if you wish)

YOUR RESEARCH DATA

Q5. How would you describe the nature of your research:

TICK ONE ONLY

○ 1 Primarily computational

○ 2 Primarily experimental

○ 3 Primarily observational

○ 4 Other (please specify)

Q6. Please select any of the following that characterise your research activity at UCL:

TICK ONE ONLY PER ROW

	Most of the time	Sometimes/Occasionally	Rarely	N/A
It is externally funded	☐ 1	☐ 2	☐ 3	☐ 4
It is internally funded	☐ 1	☐ 2	☐ 3	☐ 4
It involves collaboration with other departments/centres/institutes outside UCL	☐ 1	☐ 2	☐ 3	☐ 4
It involves collaboration with other departments/centres/institutes within UCL	☐ 1	☐ 2	☐ 3	☐ 4
It requires access to third party data	☐ 1	☐ 2	☐ 3	☐ 4
I contribute data to third party databanks	☐ 1	☐ 2	☐ 3	☐ 4
It involves dealing with what is referred to as sensitive data (e.g. clinical data)	☐ 1	☐ 2	☐ 3	☐ 4
Other (please specify)	☐ 1	☐ 2	☐ 3	☐ 4

Q7. Which of the following data types best describe your *primary research data*:

(Definition: By primary research data we refer to data produced within the timeframe of a project/research work/lifetime. Characteristics of primary research data are: unprocessed, original, generated by machines or humans):
TICK ALL THAT APPLY

- ☐ 1 Algorithm(s)
- ☐ 1 Audio(s)
- ☐ 1 Database(s)
- ☐ 1 Image(s)
- ☐ 1 Numeric data
- ☐ 1 Text
- ☐ 1 Video(s)
- ☐ 1 Other (please specify)

Q8. To what extend do the following statements apply to your research data? (By research data we refer to *any* data (primary, processed, analysed) at pre-publication stage. Please select all that apply)
TICK ONE ONLY PER ROW

	True	True to some extent	False	I am not sure/not applicable to my data
My research data are meaningful mainly in the context of my project	☐ 1	☐ 2	☐ 3	☐ 4
My research data could be useful to others outside my project	☐ 1	☐ 2	☐ 3	☐ 4
Access to my data is limited by copyright and/or other restrictions	☐ 1	☐ 2	☐ 3	☐ 4
My research data could be freely accessible for educational purposes	☐ 1	☐ 2	☐ 3	☐ 4
My data should be retained for a designated period after the end of the project	☐ 1	☐ 2	☐ 3	☐ 4
My research data can be disposed straight after the end of the project	☐ 1	☐ 2	☐ 3	☐ 4
Production of my research data involves use of GRID technology	☐ 1	☐ 2	☐ 3	☐ 4
My research data require large storage capacity	☐ 1	☐ 2	☐ 3	☐ 4
The generation of my research data is limited by availability of resources (e.g. by project's budget)	☐ 1	☐ 2	☐ 3	☐ 4
My research data are stored in means that could quickly become obsolete	☐ 1	☐ 2	☐ 3	☐ 4
My research data adhere to some disaster recovery measures (e.g. multiple backups, stored at different locations, etc.)	☐ 1	☐ 2	☐ 3	☐ 4

Q9. As an example, can you please provide the following information about any one of your research data assets? (By research data assets we mean *any* data (primary, processed, analysed) at pre-publication stage)
TICK ALL THAT APPLY

☐ 1 Name of research data asset

☐ 1 Description of the research data asset (brief description of content , purpose, etc.)

☐ 1 Owner(s) (e.g. specific researcher, University, funders, etc.)

☐ 1 Location of storage (where are the data stored/kept, e.g. network drive, CD, etc.)

☐ 1 Classification (In terms of usefulness to UCL - Minor, Important, Vital)

☐ 1 Classification comments

☐ 1 General comments

OTHER COMMENTS

Q10. We would like to discuss further issues about the provenance, ownership, location, retention and management of research assets such as the one you described in the previous section. May we contact you to arrange a short interview to discuss such issues at a time of your convenience between October-December?
TICK ONE ONLY

○ 1 Yes, I am happy to arrange a short interview (please let us have your preferred contact details)

○ 2 No, I am afraid I do not have the time during your proposed timeframe

Q11. Any other comments:
WRITE IN

Q12. If you wish to be included in the prize draw please let us have your preferred contact details:
WRITE IN

Appendix B: Information about the participating departments

Department/Centre	Number of publications[7]	Presence in UCL Eprints[8]	UCL Eprints items by year	Submitted items to RAE 2008[9]
Institute of Archaeology (Social & Historical Sciences)	2009: 3 2008:114 2007:204 2006:231 2005:192 2004:193 2003:251 2002:174 2001:119 2000:258 1999:122 1998:114 1997:94 Total = 2069	Articles:97 Books: 54 Book chapters: 82 Conference item: 1 Monograph: 5 Proceedings papers: 18 Theses: 1 Total = 258 (excludes contributions to other departments/centres) Full text in the Eprints = 13	2009: 8 2008: 22 2007: 69 2006: 48 2005: 36 2004: 28 2003: 51 2002: 28 2001: 23 2000: 2 1999: 0 1998: 0 1997: 0	Articles: 94 Books: 47 Book chapters: 71 Monograph: 3 Proceedings papers: 18 Number of items submitted: 233 Records in the Eprints = 233 Full text in the Eprints = 7
Department of Scandinavian Studies (Arts & Humanities)	2009: 0 2008:8 2007:23 2006:25 2005:12 2004:30 2003:20 2002:14 2001:19 2000:15 1999:18 1998:16	Articles:17 Books: 6 Book chapters: 8 Conference item: 1 Proceedings papers: 2 Theses: 2	2009: 0 2008: 0 2007: 9 2006: 8 2005: 5 2004: 6 2003: 4 2002: 1 2001: 3 2000: 0 1999: 0 1998: 0	Article: 16 Book: 5 Book chapter: 6 Proceedings paper: 1

[7] These numbers are based on information available on the UCL Research Publications database (http:// http://www.ucl.ac.uk/research/publications/). The information on this database is updated regularly by UCL staff (academic, research, other). The numbers listed above do not necessary reflect the whole of UCL publications as a) not all members of staff have uploaded their publications and b) not all items listed on the database are peer reviewed published articles. For example, presentations and news items are also included. Information valid on 18/02/2009
[8] UCL Eprints – http://eprints.ucl.ac.uk/ Information valid on 18/02/2009
[9] This information is hosted at the UCL Eprints: http://eprints.ucl.ac.uk/rae2008.html

	1997:22 Total = 222	 Total = 36 (excludes contributions to other departments/centres)	1997: 0	 Number of items submitted: 28 Records in the Eprints = 28 Full text in the Eprints = 2
UCL Interaction Centre (Interdepartmental and cross-faculty research)	2009: 1 2008:14 2007:14 2006:20 2005:8 2004:10 2003:11 2002:40 2001:5 2000:1 1999:1 1998:0 1997:0 Total = 125	Articles: 35 Book chapters: 6 Conference item: 5 Proceedings papers: 27 Software: 1 Theses: 2 Total = 76 (excludes contributions to other departments/centres) Full text in the Eprints = 50	2009: 1 2008: 13 2007: 14 2006: 13 2005: 8 2004: 9 2003: 5 2002: 7 2001: 3 2000: 1 1999: 0 1998: 0 1997: 0	n/a[10]
Physics and Astronomy (Mathematical & Physical Sciences)	2009: 0 2008:53 2007:140 2006:226 2005:311 2004:331 2003:360 2002:400 2001:343 2000:355 1999:280 1998:332 1997:283 Total = 3414	Articles:390 Books: 2 Book chapters: 5 Conference item: 1 Proceedings papers: 2 Theses: 2 Total = 402 (excludes contributions to other departments/centres)	2009: 0 2008: 8 2007: 41 2006: 64 2005: 63 2004: 60 2003: 62 2002: 71 2001: 32 2000: 0 1999: 0 1998: 1 1997: 0	Article: 427 Book chapter: 1 Proceedings paper: 2 Number of items submitted: 430

[10] At the moment it is not possible to identify items submitted to RAE 2008 at research centre department (due to the interdisciplinary nature of the centre UCLIC members of staff are based at more than one departments and at different faculties; therefore, would be listed under different units of assessment).

		Full text in the Eprints = 247		Records in the Eprints = 430 Full text in the Eprints = 202
Language and Communication Studies (Life & Medical Sciences)	2009: 1 2008:8 2007:7 2006:8 2005:9 2004:2 2003:12 2002:7 2001:13 2000:11 1999:4 1998:6 1997:3 Total = 91	Articles: 60 Books: 3 Book chapters: 4 Total = 67 (excludes contributions to other departments/centres) Full text in the Eprints = 3	2009: 0 2008: 1 2007: 13 2006: 12 2005: 10 2004: 5 2003: 13 2002: 11 2001: 2 2000: 0 1999: 0 1998: 0 1997: 0	n/a[11]
Speech, Hearing and Phonetic Sciences (Life & Medical Sciences)	2009: 3 2008:21 2007:32 2006:29 2005:29 2004:21 2003:27 2002:27 2001:15 2000:19 1999:34 1998:25 1997:7 Total = 289	Articles: 63 Books:5 Book chapters: 10 Conference items: 0 Proceedings papers: 6	2009: 0 2008: 0 2007: 15 2006: 15 2005: 11 2004: 17 2003: 11 2002: 12 2001: 4 2000: 0 1999: 0 1998: 0 1997: 0	n/a[12]

[11] The publications submitted to RAE2008 for the Research Department of Language and Communication Studies comes under the unit of Assessment "Allied Health Professions and Studies" which includes publications from more than one departments. Therefore, it is not feasible to provide information about the publications submitted to RAE2008 and hosted at UCL Eprints.
[12] ibid

		Total = 85 (excludes contributions to other departments/centres) Full text in the Eprints = 3		

Department/Centre	Research field	Number of academic and research staff contacted	Response to the questionnaire survey	Interviews	Type
Institute of Archaeology	Social & Historical Sciences	71	25	13	Only permanent members of staff, no PhD students
Department of Scandinavian Studies	Arts & Humanities	10	-	6	Current and honorary members of staff; no PhD students
UCL Interaction Centre	Interdepartmental and cross-faculty research	14	5	3	Only current members of staff; no Ph.D. students
Physics and Astronomy	Mathematical & Physical Sciences	68	13	4	Only current members of staff; no PhD students
Language and Communication Studies	Life & Medical Sciences	17	6	3	Only current members of staff; Yes - PhD students
Speech, Hearing and Phonetic Sciences	Life & Medical Sciences	22	8	1	Only current members of staff; no PhD students

Department/Centre	Research field	Number of academic and research staff by faculty [13]	% of members of staff by faculty [14]
Institute of Archaeology	Social & Historical Sciences	324	8.1
Department of Scandinavian Studies	Arts & Humanities	228	5.7
UCL Interaction Centre	Interdepartmental and cross-faculty research	n/a	n/a
Physics and Astronomy	Mathematical & Physical Sciences	478	11.9
Research Departments of : Language and Communication Studies and Speech, Hearing and Phonetic Sciences	Life & Medical Sciences	595	14.9

[13] Data available at the UCL Research website: http://www.ucl.ac.uk/research/facts
[14] The number of staff by faculty for all faculties is available at http://www.ucl.ac.uk/research/facts. In this table we only list the number of staff by audited department.

Appendix C: Reaction to DAF

Purpose – Scope related

"The purpose and value of this exercise remains completely opaque, and I haven't found the information on the web-sites below[15] helpful in explaining either of these (some links lead to unfinished pages; flow diagrams convey nothing of substance). I have not read the pilot reports provided there, as a quick glance suggested they have been the type of projects which are designed principally to keep those doing them employed; useful outcomes for the research community are not apparent or clearly stated, and I simply do not have the time to fight through a lengthy report in bureaucratic jargon to find out its point. Unless you can direct me to something indicating why this isn't simply another top-down waste of time, a further hour invested in it has little attraction".

Social & Historical Sciences.

Classification of data (minor, important, vital)

- *No idea..... what does UCL regard as vital?*
- *I don't understand this question*
- *Not sure what this means*
- *Important for analysis in future publications*
- *Vital for our research*
- *useless to UCL; vital to me if I want to publish*
- *I am not sure*

Various from the questionnaire survey

Research data in general

"Not at all, it was interesting to have this opportunity to think about my research in ways I don't normally, and I am glad to be of some help"

Humanities

Forms – Metadata

[15] The reference is to the http://www.data-audit.eu webpage

“Sorry to be a bore, but I can't make head or tail of the Excel file you sent me. It bears no relation to the Audit form 3A you gave me in hard copy, and I therefore have no idea what to do with it. I tried going into www.data-audit.eu, but that seemed to require me to download some huge document first with no guarantee of what I would find at the end of it. I was reluctant to do this because I don't want to clog up my computer with unnecessary data. If you can send me something that is a replica of Audit Form 3A, I will do my best with it. Otherwise I am stuck. But maybe my incomprehension at this stage provides you with useful information about workers in the humanities anyway?”

Humanities

“Immediately…most researchers would take a look at it and think “bloody hell”…come on, what else guys?!

Social & Historical Sciences

“Filling in the form for Particle Physics could be difficult – we generate Petabyes of data per year in many, many different formats and itemising how we deal with each of them could be quite time consuming and potentially not that interesting to you”.

Mathematical & Physical Sciences