



RESEARCH DATA MANAGEMENT SERVICES: FINDINGS OF THE CONSULTATION WITH SERVICE PROVIDERS

SCOPING DIGITAL REPOSITORY SERVICES FOR RESEARCH DATA MANAGEMENT

www.ict.ox.ac.uk/odit/projects/digitalrepository/

Author	Luis Martinez-Urbe (luis.martinez-uribe@oerc.ox.ac.uk) Digital Repositories Research Co-ordinator
Document name	ConsultationWithOxfordServiceProviders.doc
Version	2.1
Date created	19/9/08 15:52
Date last modified	Last saved by Luis Martinez-Urbe on 27/3/09 14:40
Distributed to	Prof. Paul Jeffreys, Richard Ovenden, Dr. Mike Fraser and members of the Oxford Digital Repositories Steering Group.

A collaborative project between

OFFICE OF THE DIRECTOR OF IT
Enabling Oxford University to make optimal use of IT



Oxford University Library Services

Oxford Digital Repositories Steering Group

Contents

EXECUTIVE SUMMARY	2
1. METHODOLOGY	5
1.1 Scope	5
1.2 Identification of candidates to participate	5
1.3 Organization of consultation interviews	5
1.4 The workshop.....	5
2. OXFORD UNIVERSITY SERVICE PROVIDERS	6
2.1 Library Services.....	6
Oxford Research Archive.....	6
Oxford Digital Library	6
Social Science Library.....	6
2.2 Computing Services.....	7
Hierarchical File Server.....	7
Research Technologies Services.....	7
NSMS	7
IT Learning Programme	7
2.3 Oxford e-Research Centre	7
Oxford Supercomputing Centre.....	8
National Grid Service	8
OxGrid	8
2.4 Research Services.....	8
2.5 Other Services	8
Information Management Services Unit.....	8
Social Sciences Data Library – Nuffield College.....	9
Software Engineering Research Group - Computing Laboratory	9
ISIS Innovation	9
Central University Research Ethics Committee	9
The Doctoral Training Centre	9
Medical Sciences Skills Training Programme	9
Centre for Research Methods in Social Sciences.....	10
Divisional Research Facilitators/Coordinators	10
3. THE INSTITUTIONAL AND NATIONAL SERVICES FOR RESEARCH DATA MANAGEMENT WORKSHOP.....	11
4. RESEARCH DATA MANAGEMENT AND CURATION SERVICES FRAMEWORK.....	14
4.1 From requirements to a set of services.....	14
4.2 Mapping the DCC Lifecycle Model and the Research Data Management and Curation Services Framework	15
5. DISCUSSION OF RESULTS.....	17
6. NEXT PRIORITIES.....	20
7. RECOMMENDATIONS.....	20
APPENDIX 1. FRAMEWORK FOR CONSULTATION WITH SERVICE PROVIDERS	21
APPENDIX 2 – RESEARCH DATA MANAGEMENT SERVICES DIAGRAM	22
APPENDIX 3 – WORKSHOP INSTITUTIONAL AND NATIONAL RESEARCH DATA MANAGEMENT SERVICES	23

EXECUTIVE SUMMARY

The Scoping Digital Repository Services for Research Data Management project is a cross-agency collaboration in Oxford. This report presents the findings of the consultation with Oxford service units, the complementary workshop organized to hear about data management services and the recommendations made by the Oxford Digital Repositories Steering Group for continuing work in this area.

A total of eleven service units across the university took part in the consultation exercise to validate the requirements captured from researchers earlier on in the project and to help defining the services on offer, or planned to be offered, that support researchers with their data management duties.

A framework for research data management and curation services is presented below, this was derived from the researchers' requirements for services and complemented with some of the stages present in the DCC Curation Lifecycle Model and the feedback provided by service units in Oxford as well members of the UK Research Data Service and the Digital Curation Centre.

Research Data Management and Curation Services Framework												
DATA MANAGEMENT & SHARING PLANS	LEGAL & ETHICAL	BEST FORMATS & BEST PRACTICE	SECURE STORAGE	METADATA	ACCESS & DISCOVERY	COMPUTATION ANALYSIS & VISUALIZATION	RESTRICTED SHARING	DATA CLEANING	PUBLICATION	ASSESS VALUE	PRESERVATION	ADD VALUE
TOOLS & INFRASTRUCTURE												
ADVICE & SUPPORT												
Data Management and Sharing Plans	Support and advice to help researchers prepare their data management and sharing plans.											
Legal and Ethical	This service includes support to assist researchers with the legal and ethical implications of creating, sharing and using data.											
Best Formats and Best Practice	Support for researchers to decide which are the best formats and practice for producing and documenting specific data. This service may also include provision of support for database design.											
Secure Storage	Secure storage includes infrastructure that allows storing research data providing backup and version control capabilities amongst other things.											
Metadata	Tools and support to permit researchers describe their data from the moment of creation.											
Access and Discovery	A support service as well as tools to help researchers locate research data and access research data. This service could also include tools to help research groups to find about their data resources using the Data Audit Framework methodology.											
Computation, Analysis & Visualization	Software and computing resources that allow analysis and visualization of research data as well as the training needed to equip researchers with the appropriate skills.											
Restricted Sharing	Technical infrastructure to share research data with selected individuals or groups.											
Data Cleaning	Support to clean and prepare data to the standard required for publication. This service should include help with anonymizing data.											
Publication	Infrastructure that permits researchers to publish documented data and link them to research articles and other materials located in other repositories. In some cases researchers may want to exploit their data commercially. DRAMBORA could serve as a											

	tool here to assess repositories that publish the data.
Assess Value	One of the main challenges with research data is deciding what data needs to be kept and for how long.
Preservation	This service would be responsible for looking after the data in the long-term applying the required measures so that the data is accessible through time.
Add Value	Once the data is stored with the metadata associated with it, value can be added by organizing similar data in groups, promoting it, linking it to other materials or allowing annotations.

When examining the level of service provision in Oxford for the different research data management services proposed and organizing those into three levels of service, see figure below, it is apparent that most services from the framework are not fully available to researchers.

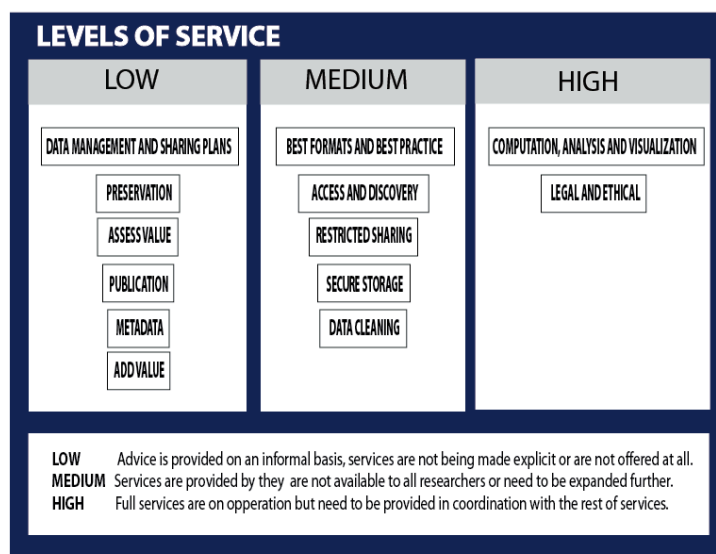


Figure 2. Level of service provision for the research data management services

Expertise in the different areas of data management and curation is widespread amongst service units at Oxford. In many cases support is provided in an individual and ad-hoc basis but services are not being made explicit as per the lack of resources to offer those.

On the whole, the vast majority of the research data management services identified are not being offered fully or at all by service units across the University. Moreover, those services currently on offer would benefit of coordination to make them more seamless to the researchers using them.

In addition to this, there is a need for a University wide policy for management and curation of research data as well as for provision of advice and guidance to service units at all levels.

The findings from this consultation were presented to the Oxford Digital Repositories Steering Group (ODSRG) in the November 2008 meeting and the following two main recommendations were produced:

- I. 1 FTE post should be created to be proactive in raising awareness and provide support for research data management;**
- II. Infrastructure services should be developed to support the research data lifecycle recognizing that resources already exist.**

INTRODUCTION

The project Scoping Digital Repository Services for Research Data Management started in January 2008 as a collaborative effort between the Oxford University Computing Services, the Library Services, the e-Research Centre, the Office of the Director of IT and the Oxford Digital Repositories Steering Group. The aim of the project is to scope the requirements for digital repository services to manage and curate research data.

The initial main activity of the project involved capturing researcher's requirements for services to help them manage their data more effectively. To do this, interviews were conducted with researchers across disciplines and a complementary workshop was organized. Both, the interviews and the workshop aimed to document researchers' practice with data in addition to capture their requirements. This effort formed also the basis for the case study conducted in Oxford as part of the UK Research Data Service (UKRDS) feasibility study. Results from this exercise revealed the variety of practices and diverse maturity levels in terms of research data management. It was also clear from the exceptional participation in the interviews and the workshop, that many researchers and research groups have an interest in the topic as they are facing the challenges of managing their research data at present.

Once researchers' requirements for services are better understood, the next step is to comprehend the services provided by the different service units in the university and map services available to those required by researchers. By doing this it is possible to identify the gaps in service provision as well to stimulate discussion amongst service units on how to best coordinate efforts to provide effective services to meet the needs of researchers at Oxford.

This report describes the activities and findings of the consultation with services providers in Oxford. It aims to inform the Oxford Digital Repositories Steering Group as well as service units across Oxford about the research-driven requirements for repository services to manage and curate research data and how service units across the Collegiate University serve the needs of their researchers.

The report is organized as follows: the methodology used for the study is described; each of the services provided by the units that participated in the study is briefly presented. The report includes the findings from a complementary workshop organized in October 2008. A framework of research data management services deduced from researchers' requirements and feedback from service units is presented and mapped to the DCC Curation Lifecycle Model. To finish, the results of the consultation exercise are discussed and the next priorities of the project are outlined.

1. METHODOLOGY

In this section the methodology followed to conduct the consultation with service providers in Oxford is explained. This methodology mirrors the one used for gathering the researchers' requirements in the first part of the project¹, which was largely based on the methodologies used by the Building a Virtual Research Environment for the Humanities (BVREH)² and the e-Infrastructure Use Cases³ projects.

1.1 Scope

The consultation with service providers in Oxford aimed to validate the researchers' requirements for services gathered through the scoping study interviews as well as to determine the data management services available to researchers and plans for future ones. A series of interviews with service providers and a workshop were arranged and complemented with some desk research for this consultation exercise. The focus was primarily on Oxford central support services and the data management services they provide or plan to provide in the next 3 to 5 years.

1.2 Identification of candidates to participate

In order to identify suitable candidates to participate in the consultation exercise a combination of several approaches was adopted. Choice of candidates was originally guided by suggestions from the internal management of the project, for initial group of candidates. For each member of staff participating, the friend of a friend approach was used to identify more candidates.

1.3 Organization of consultation interviews

The potential participants identified were emailed individually providing them with the project brief and asking them to take part on the consultation. If they agreed, the interview questions were circulated and time and place was arranged for the meetings to take place.

In general interviews took no longer than one hour and required one hour of preparation to learn more about the support services where participants work. The interview itself (see appendix 1) started with a brief introduction to the project and a reminder of the nature of the consultation as well as the intention to take notes. The aim of the interviews is to validate the requirements for services gathered in the scoping study interviews, learn about those services on offer as well those planned to be developed in the next 3 to 5 years. During the interview the Research Data Management Services Diagram (see appendix 2) was used as a visual tool to show participants the range of services researchers require.

1.4 The workshop

The workshop was originally planned to bring speakers from national facilities for data management and other institutions to raise awareness of the services already on offer elsewhere that could potentially be delivered in Oxford. In addition to this, a panel session with representatives from the different service providers in Oxford was organized to stimulate discussion about the roles they could play to support researchers with their data management. See appendix 3 for the full programme.

¹ Luis Martinez Uribe (2008) "Findings of the Scoping Study and Research Data Management Workshop" available at: <http://tinyurl.com/55fxgw>

² Building a VRE for Humanities website available at: <http://bvreh.humanities.ox.ac.uk/>

³ e-Infrastructure Use Cases website available at: <http://www.eius.ac.uk/>

2. OXFORD UNIVERSITY SERVICE PROVIDERS

In this section, a selection of services related to data management activities provided by a variety of service units in Oxford are described. Due to the devolved structure in the University of Oxford, service units are available centrally like Computing Services or Library Services and also at the more local level in colleges and departments. Therefore it is highly complex to present a complete picture of service providers and their services at all levels. In this report, the focus is in the central service units that serve all the university or at least one division.

2.1 Library Services

The Oxford University Library Services (OULS) is an integrated library service for the University with the mission of responding to users' library needs and maintaining and developing access to Oxford's collections⁴. Here the focal point is a selection of some OULS units and services that relate closely to the management and curation of digital data resources and collections.

Oxford Research Archive

The Oxford University Research Archive (ORA) provides secure storage, management, access and preservation to research outputs produced by members of the University of Oxford. Mostly those outputs include journal articles, conference papers and theses but recently they have started dealing with the archiving and curation of some small datasets⁵ as well as referencing large datasets⁶ hosted by the Oxford e-Research Centre. Although members of ORA offer advice occasionally on best formats for data creation or copyright issues, these services are not being made explicit in many cases as there are not enough resources to fully offer them.

Oxford Digital Library

The Oxford Digital Library (ODL) is a service to support the creation of digital materials from the University Library holdings. Their services include imaging capture tools, advice and support to projects that involve the creation, description, storage and dissemination of digital materials. ODL also hosts a number of digital collections that are mainly focused in the research areas of Social Sciences and Humanities and provides digital preservation for those collections.

Social Science Library

The Social Science Library (SSL) in Oxford provides services to support Social Science scholars. Amongst those services they have particular expertise in data resources. The SSL subscribes and provides web-access to a number of data resources of interest to social scientists. In addition to this, tools and support are supplied to help researchers discover data. Moreover, the SSL is involved in the EU funded project⁷ that has a work-package aimed at ingest into ORA economic datasets produced by researchers.

⁴ Oxford University Library Services (2006) "Vision for 2011" available at: <http://www.ouls.ox.ac.uk/docs/vision2011.pdf>

⁵ As part of the DataShare project, website available at: www.disc-uk.org/datashare.html

⁶ As part of the BID project, website available at: <http://www.jisc.ac.uk/whatwedo/programmes/reppres/sue/bid.aspx>

⁷ The NEEO project, website available at: <http://www.nereus4economics.info>

2.2 Computing Services

Oxford University Computing Services (OUCS) is the central ICT service provider and responsible for services that are direct to the user such as networks, backup or email⁸. Below some of the services that are relevant to this study are explained.

Hierarchical File Server

The Hierarchical File Server (HFS) is a back up service and secure long-term storage available, amongst others, to researchers and research groups in Oxford. The back-up service is aimed at securing active files from desktops, servers and departmental servers. The long-term archive is aimed for data that is considered to be of value to the University. Since data cannot be archived indefinitely, projects are allocated storage for between one to seven years. Long-term projects and those requiring large amounts of data storage (above 1Tb) may be asked to contribute to the costs. A new HFS policy has been developed where these costs are explained. This policy also makes clear that HFS provides the long-term file storage and not a data curation service. Consequently it defines the role of a Data Curator, every dataset must have one, responsible for submitting the data, ensuring it is documented and reviewing it on a regular basis.

Research Technologies Services

The Research Technologies Service (RTS) acts as point of access to the different OUCS services for the research community. Whilst the Service is intended for all academic divisions, there is significant usage from within the Humanities Division (where the level of local IT support may be less than within other divisions). The RTS comprises advice and support for the creation and description of datasets, especially textual data. The Oxford Text Archive provides storage, description, access and preservation for textual data.

NSMS

The NSMS service offers a wide range of chargeable services to departments in the University. The ones of interest to this study include server management which could potentially be used as a shared file storage facility and web hosting which again could potentially be used to publish research data on the web.

IT Learning Programme

The IT Learning Programme (ITLP) provides a wide range of taught courses including design, structure and querying databases, two copyright courses for creating or using online materials and statistical packages such as SPSS, STATA, MapInfo, Matlab or NVivo.

2.3 Oxford e-Research Centre

The Oxford e-Research Centre (OeRC) collaborates with other units in the University to facilitate interdisciplinary and collaborative research using innovative information and communication technologies. Apart from the OeRC services explained below, there are a number of projects that are of relevance to this study such as the BVREH (Humanities) and Bridging the Interoperability Divide (with OULS and OUCS). In addition to this, there are plans to provide visualization services in the future.

⁸ Paul Jeffreys (2008) "Oxford's Computing Model: Delivering Computing Services to the University" available at <http://www.ict.ox.ac.uk/odit/ITcoordination/ODITOxford%27sComputingModel.pdf>

Oxford Supercomputing Centre

The Oxford Supercomputing Centre (OSC) provides high-performance-computing (HPC) services for Oxford researchers. This service comprises the technical infrastructure, clusters and shared memory systems, to speed up to 30 times research work carried out on desktop computers. In addition to this, they provide advice and support, including courses, on scientific computing ranging from application to programming.

National Grid Service

The National Grid Service (NGS) is a national consortium made up of four core partner institutions and fourteen other partner and affiliate institutions. NGS delivers grid services including access to computational resources, storage, data and applications for cross-institutional collaborations. Although storage space is provided the long-term cannot be warranty, a data storage service (the Storage Resource Broker) for data in distributed locations is provided, data can be shared in the NGS in a controlled way by using grid security mechanisms and there is support available for database design.

OxGrid

OxGrid is Oxford's infrastructure to provide University researchers seamless access to computing and storage resources available in departments and colleges. This service integrates with the National Grid Service and the Oxford Supercomputing Centre.

2.4 Research Services

The Research Services Office (RSO) support the research grant process by providing advice on funding opportunities, reviewing grant applications and promoting the responsible conduct of research and compliance with regulatory and sponsor requirements and the continuous improvement of research administration at Oxford. RSO is first point of contact with researchers and they can also refer to other services and expertise available in the University.

2.5 Other Services

The services presented below represent a selection of units that provide some advice, guidance or training related to data management.

Information Management Services Unit

The Information Management Services Unit (IMSU) serves around 3500 users across the medical sciences division. IMSU provides secure data storage services, currently hold 21Tbs, for its users with mirror copies and back ups using the HFS. The storage service is not suitable for confidential data. IMSU also support users who want to share data with collaborators by setting up FTP servers on request. There is also provision of advice and support on technical aspects related to data storage. Annotation, discovery and access are seen to be the user's responsibility. A certain level of legal advice is also provided but they refer users to University lawyers for complex legal queries. They have in the past advised their users on data preservation but see this is as a role falling out of their remit.

Social Sciences Data Library – Nuffield College

The Social Sciences Data Library has been running since 1988. The Data Manager responsible for the service assist researchers from the Social Sciences Division to locate and format data, negotiate access arrangements and provides support for questionnaire design, data collection and data cleaning methods. In addition to this, the Data Library hosts a collection of survey micro-dataset from UK and elsewhere as well as subsets of key variables of government surveys combined over time.

Software Engineering Research Group - Computing Laboratory

The Computing Laboratory does not generally provide services per se. Nevertheless, through the Generic Infrastructure for Medical Informatics (GIMI) project⁹, researchers have developed technologies to support the secure sharing of research data. Through a standard interface that supports fine-grained access control policies, researchers can share their data with collaborators or aggregate data from disparate sources. The system currently supports a variety of applications, including: the MRC-funded NeuroGrid project, a prototype demonstrator for the National Cancer Research Institute, a demonstration application for the Health Information Research Institute, and a variety of applications within GIMI (image analysis for cancer care; self-management of long-term conditions; radiologist training).

ISIS Innovation

ISIS innovation is Oxford technology transfer company managing the University's intellectual property portfolio. Marketing research output is one of the services that ISIS Innovation offers to University researchers and consequently they advise researchers on how to best commercialise data produced as part of their research activities.

Central University Research Ethics Committee

The Central University Research Ethics Committee (CUREC) provides an ethical review process for research involving human participants for those research disciplines without the appropriate machinery. CUREC provides guides for data collection and storage where issues such as data ownership, anonymity as well as retention and disposal of data are explained. There is likely to be a requirement from CUREC to record those research projects where ongoing monitoring of ethical conduct is a requirement of an external organisation providing funding for the research, there are already at least two examples of this (ESRC and NIH).

The Doctoral Training Centre

The Doctoral Training Centre (DTC) delivers training for doctoral researchers in physical and life sciences in Oxford. In addition to providing the basic background in biology, biochemistry and biological physics, the DTC aims to provide more practical research skills in mathematical modelling, scientific computing, computer programming and statistical methods. At the moment there are not courses related to the creation, management and curation of data although it is intended to include courses of this type in the future.

Medical Sciences Skills Training Programme

The Skills Training Programme in Medical Sciences trains research students and post-doc research assistance on a variety of areas. The training courses delivered include Data Protection and statistical training with SPSS.

⁹ The GIMI project website is available at: <http://www.gimi.ox.ac.uk/>

Centre for Research Methods in Social Sciences

The Centre for Research Methods in Social Sciences (ReMiSS) provides teaching and advice to students and researchers in all Social Sciences in Oxford, except the Department of Economics, on the application of research techniques and methods. Some of the courses taught include applied statistics and statistical methods using real datasets and statistical packages such as SPSS, STATA or R.

Divisional Research Facilitators/Coordinators

The divisional Research Facilitators/Co-ordinators support researchers in their divisions to identify funding opportunities and to produce complete and competitive applications. Research Facilitators/Coordinators manage the internal and external process of application and also act as gateways between researchers and support services available elsewhere. When funding bodies require data management plans to be included in applications, researchers will be referred to appropriate personnel within the division and outside of it.

3. The Institutional and National Services for Research Data Management Workshop

The second workshop organized as part of the scoping study was held on 20th October at the Said Business School (see appendix 3 for complete programme) to complement the consultation with service units in Oxford. The event brought around 45 delegates from the Oxford departments and units shown below as well as representatives from the Digital Curation Centre, the UK Research Data Service and the Joint Information Systems Committee (JISC).

Departments/Colleges and other units	
Begbroke Directorate	Oxford e-Research Centre
Chemistry Research Laboratory	Oxford Institute of Ageing
Computing Laboratory	Oxford Internet Institute
Computing Services	Oxford University Library Services
Image Bioinformatics Research Group, Department of Zoology	Radiation Oncology and Biology
Institute of Archaeology	Research Services
Legal Services	Saïd Business School
Nuffield College	Sociology
Office of the Director of IT	Wellcome Trust Centre for Human Genetics

Table 1. Affiliations of those attending the workshop

Professor Paul Jeffreys provided the introductory speech where he highlighted the need for collaboration amongst service units in Oxford to solve one of the main challenges universities are facing, that of supporting their researchers to manage their data. After this, Dr. Natasha Balac, from San Diego Super Computer Center (SDSC), introduced the SDSC Data Central service, an innovative unit within SDSC that provides expertise, infrastructure and tools for data management. It was originally funded by the National Science Foundation and currently hosts around 130 data collections including some of great importance such as the Protein Data Bank. Their integrated infrastructure is made of four main components: data storage, data management, data manipulation and data access. This infrastructure then translates to services that include:

- Database hosting;
- Long-term storage and preservation;
- Remote data management and access;
- Data Analysis, visualization and data mining;
- Professional support.

Some of the projects Natasha mentioned that SDSC Data Central is involved with included TeraGrid¹⁰ and Chronopolis¹¹. TeraGrid integrates high-performance computers, data resources and tools, and high-end experimental facilities from eleven partner institutions around the US. This is done through *science gateways* that allow use of resources such as data, applications or storage through a common interface. Chronopolis is a digital preservation demonstration project to develop a framework for preservation grids. Organizations such as the Library of Congress, the Inter-University Consortium for Political and Social Research or the California Digital Library are providing digital collections to the Chronopolis preservation grid which creates geographically distributed copies of those, while enabling curatorial audit reporting and access for preservation clients.

¹⁰ The TeraGrid project website is available at <http://www.teragrid.org/>

¹¹ The Chronopolis project website is available at: <http://chronopolis.sdsc.edu/>

Neil Beagrie (Charles Beagrie Ltd.) reported on the UK Research Data Service (UKRDS) feasibility study. After the work carried out in the four case study institutions to capture the researcher's viewpoint, the UKRDS team proposed an hybrid/umbrella model representing the interests of both the national facilities and the emerging institutional repositories. A final report for the feasibility study will be completed by December.

The Data Audit Framework (DAF) and the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) are two tools/methodologies developed by the Digital Curation Centre and Sarah Jones introduced them in the workshop. The DAF provides a method to assess what data are stored in departments or research centres and how they are being managed. DRAMBORA is a toolkit for repository managers to self-assess their capabilities.

Louise Corti, from the UK Data Archive, talked about the importance of sharing data and the significance of the work in this area from the Economic and Social Research Council, with a data policy since 1995. Then she described the data support service established for the interdisciplinary Rural Economy and Land Use (RELU) project¹² to help researchers manage the data created as part of the project. Louise highlighted throughout her talk the need to educate and support researchers to ensure that data are managed appropriately from the moment of creation.

Dr. Stuart Jeffrey's presentation described the services provided by the Archaeology Data Service (ADS) around five main areas: digital preservation, data access, guides to good practice, teaching and learning resource and a technical advisory service. For digital preservation, they follow the Open Archival Information System model (OAIS), a standard for the long-term preservation of digital objects. ADS provides access to their collections and metadata online but they also allow discovery of archaeological data collections elsewhere.

From the NERC Environmental Bioinformatics Centre, Tim Booth, described how they support biological researchers using molecular tools. Their belief is:

"When researchers work closely with data management specialists throughout a project there are immediate and longer-term benefits."

And their services are arranged around four main themes: data management, bioinformatics workstations, training and support and software and standards development. Tim presented a case study of marine metagenomics and how they supported them with expertise on database design and tools like the Bio-Linux workstations.

Carolyn McKee, from Legal Services in Oxford, gave the last presentation. Carolyn pointed out at the increased risk of managing institutionally digital content and how these come from materials in digital repositories that are to be deposited and accessed. The main rights to consider include copyright, database rights, moral rights, contractual rights and confidence. She called for taking adequate licences from depositors as well as protecting against claims from users. She mentioned at the end that legal Services in Oxford supply legal help and advice to university members in issues around IP.

The panel session at the end of the workshop, facilitated by Prof. Paul Jeffrey, started the discussion about the roles and responsibilities of a selected group of services units in Oxford and the UKRDS. Sally Rumsey, from the Oxford Research Archive (ORA), explained how ORA houses Oxford research materials although its scope for research data is still not defined. Sally pointed out how ORA can take small datasets, link to publications and point to big datasets held somewhere else. She highlighted the importance of understanding that this is not only a storage

¹² The Rural Economy and Land Use, website available at <http://www.relu.ac.uk/>

problem but also digital preservation one, the need for expertise and continuing resources to deal with the curation of digital materials and how crucial it is to fit with other national and international developments.

Dr. Mike Fraser from the Computing Services in Oxford explained the services available through the Research Technologies Service to provide support and advice for data encoding as well as for funding proposals related to IT aspects. Mike also mentioned the Hierarchical File Server (HFS) and its dual role for back up and archival of data. To finish, he emphasized the need for sustainable funding to finance these activities and he spelt out his wish to establish a service similar to the SDSC Data Central that could address Oxford's requirements.

Dr. Jon Lockley from the Oxford e-research Centre and the Oxford Super Computing Centre (OSC), explained the high performance computing services provided by OSC and how the OSC was never geared up for long-term management of data. OSC's storage resources are being used as a dumping ground for big datasets because there is no other place to put them; with a consequent significant increase in the scale of storage requested by OSC users. Jon asked for a more joined-up approach to deal with researchers' needs.

Kathryn Dally from the Research Service Office (RSO) explained how the RSO acts as a focal point for research enquiries and their role for administering applications for research funding and monitoring funders policies. Kathryn pointed out the policy vacuum in this area and the need for a central policy for data and records that establishes for how long to keep the data, guidance on research with human participants and other specific domain guidance such as laboratory notebooks.

In Neil Beagrie's view, there is an excellent opportunity for UKRDS to add value by providing coherence at the national level and addressing the gaps but he stressed the need to work hard for greater coordination. When asked whether UKRDS would only add another layer of bureaucracy, Neil responded that the scale is important and not leveraging activities could cost a lot of money, besides many institutions could not afford to provide data support services.

The panel was asked about the data stored on the OSC and this started a discussion about how to best assess the value of the data, where and how to move them and whether the input files that created the simulations are more important to be preserved than the actual raw data.

Natasha Balac stressed the importance of identifying quick wins and build on those. She explained how SDSC Data Central started as a service to deal with the requirements of researchers from Life Sciences' and their simulation data and how it then expanded its remit to other disciplines.

Mike Fraser suggested a model where there is central provision of certain data management services whilst retaining devolved responsibility to the relevant local units. Louise Corti made the point that whatever infrastructures are available it is crucial to never underestimate the user support required, including outreach, promotion and training.

Overall, the day brought a wide range of experiences of data service providers and encouraged discussion about roles from service units in Oxford. Throughout the day a wide range of services were described including infrastructure and tools for storage, access, discovery, use or preservation as well as services related to support, advice and training starting as early as possible in the research lifecycle. The final panel discussion between some of the service units in Oxford evidenced the need for coordination and funding to provide the range of support services that Oxford researchers need.

4. Research Data Management and Curation Services Framework

This section presents a framework of research data management and curation services that has its origin in the requirements gathered from researchers in the first part of the project. In addition to this, a mapping is provided between the data management services and the DCC Digital Curation Lifecycle Model¹³ that defines the stages needed for curating digital material.

4.1 From requirements to a set of services

The interviews conducted between May and June helped gathering enough evidence about researcher's challenges with data. This evidence served to compile the top requirements, shown below, for services to help Oxford researchers manage their data more effectively.

Advice on practical issues related to managing data across their life cycle. This help would range from assistance in producing a data management/sharing plan; advice on best formats for data creation and options for storing and sharing data securely; to guidance on publishing and preserving these research data.

A secure and user-friendly solution that allows storage of large volume of data and sharing of these in a controlled fashion way allowing fine-grained access control mechanisms.

A sustainable infrastructure that allows publication and long-term preservation of research data for those disciplines not currently served by domain specific services such as the UK Data Archive, NERC Data Centres, European Bioinformatics Institute and others.

Funding that could help address some of the departmental challenges to manage the research data that are being produced.

The service units consulted validated the above top-level research-driven requirements. From these requirements in conjunction with the DCC Digital Curation Lifecycle Model, a set of services was generated involving advice and support as well as tools and infrastructure to support researchers in their work with data across the research lifecycle. This set of services was presented to the different service units across Oxford during the consultation. In addition to this, staff from the Digital Curation Centre and the UKRDS provided feedback about the services. After several iterations the final set of services is described next. These services should not be provided in isolation because of their many strong interdependencies.



Figure 1. Research data management and curation services framework

¹³ The DCC Digital Curation Lifecycle Model is available at: www.dcc.ac.uk/docs/publications/DCCLifecycle.pdf

Data Management and Sharing Plans	Support and advice to help researchers prepare their data management and sharing plans.
Legal and Ethical	This service includes support to assist researchers with the legal and ethical implications of creating, sharing and using data.
Best Formats and Best Practice	Support for researchers to decide which are the best formats and practice for producing and documenting specific data. This service may also include provision of support for database design.
Secure Storage	Secure storage includes infrastructure that allows storing research data providing backup and version control capabilities amongst other things.
Metadata	Tools and support to permit researchers describe their data from the moment of creation
Access and Discovery	A support service as well as tools to help researchers locate and access research data. This service could also include tools to help research groups to find about their data resources using the Data Audit Framework methodology ¹⁴ .
Computation, Analysis & Visualization	Software and computing resources that allow analysis and visualization of research data as well as the training needed to equip researchers with the appropriate skills.
Restricted Sharing	Technical infrastructure to share research data with selected individuals or groups.
Data Cleaning	Support to clean and prepare data to the standard required for publication. This service should include help with anonymizing data.
Publication	Infrastructure that permits researchers to publish documented data and link them to research articles and other materials located in other repositories. In some cases researchers may want to exploit their data commercially. DRAMBORA could serve as a tool here to assess repositories that publish the data.
Assess Value	One of the main challenges with research data is deciding what data needs to be kept and for how long.
Preservation	This service would be responsible for looking after the data in the long-term applying the required measures so that the data is accessible through time.
Add Value	Once the data is stored with the metadata associated with it, value can be added by organizing similar data in groups, promoting it, linking it to other materials or allowing annotations.

4.2 Mapping the DCC Lifecycle Model and the Research Data Management and Curation Services Framework

The DCC Curation Lifecycle Model was produced by the Digital Curation Centre to provide a high level overview of the stages required to fully curate and preserve digital material from the early point of conceptualization. As Sarah Higgins¹⁵ pointed out, a lifecycle approach enables visualization of the processes, activities and relationships. It also helps to plan activities by mapping the phases to roles and responsibilities, standards and technologies to be implemented.

Here, the sequential actions of the model are mapped, see figure 2, to the research data management and curation services that were deducted from researchers' requirements. This helps validate the framework proposed earlier on against a lifecycle model for curation of digital material. It also assists to characterize when the different services will be needed and applied during the lifecycle of the digital materials to be

¹⁴ The Data Audit Framework has been developed by the DCC and JISC, see: <http://www.data-audit.eu/>

¹⁵ Higgins, S. (2007) Draft DCC Curation Lifecycle Model. *International Journal of Digital Curation*, 2.

managed.

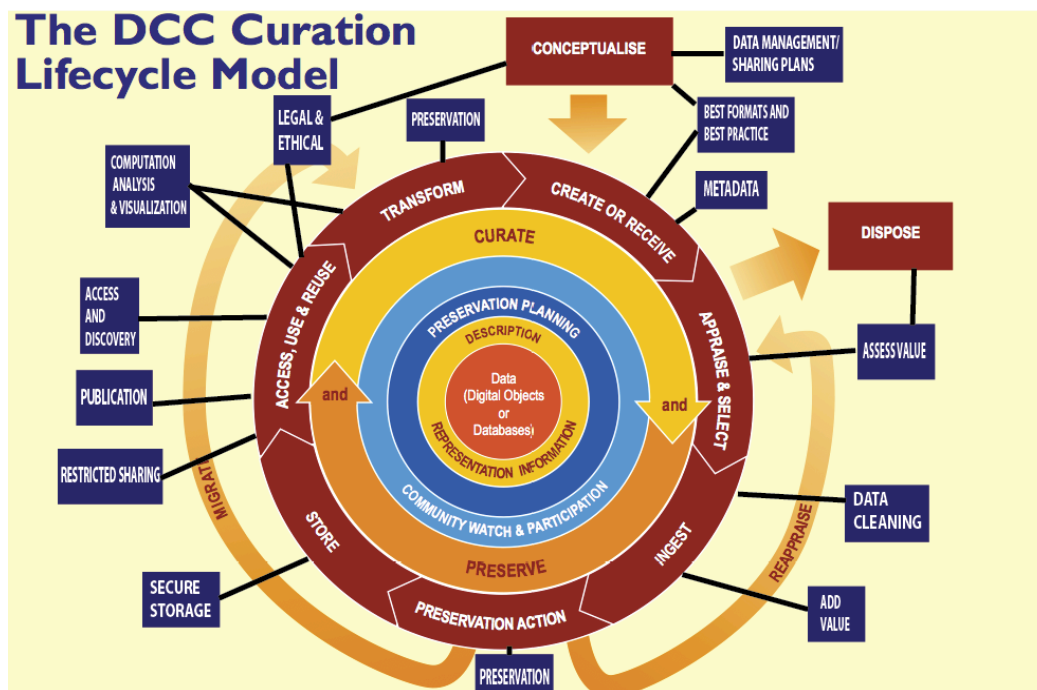


Figure 2. Research data management and curation services mapped to the DCC Curation Lifecycle Model

Mapping between DCC Curation Lifecycle and Research Data Management Services		
DCC Lifecycle Model Sequential Actions	Research Data Management Services	Description
Conceptualise	Data Management/ Sharing Plans; Best Format and Best Practice; Legal and Ethical	This stage is related to services to support researchers in the production of data management and data sharing plans. It is also related to the advisory services for best formats and best practice for data creation as well as legal and ethical services for data creation (for instance to clearly define the ownership of the data to be created or how can they be used) and sharing.
Create or receive	Best Format and Best Practice; Metadata	At this point researchers need support to figure out best formats and best practice as well how to best document their data with appropriate metadata.
Appraise and select/ dispose	Assess Value	This phase relates to services to assess value of the data.
Ingest	Data Cleaning, Add Value	Before data are ingested, they will need to be prepared and cleaned. During ingestion other information can be added to enhance them.
Preservation Action	Preservation	Obviously relates to preservation services.
Store	Secure Storage	This stage clearly relates to the secure storage.
Access, Use and Reuse	Publication; Legal and Ethical; Computation, Analysis and Visualization	This phase relates to several of the research data management services. Publication of data as well as access and discovery belong to this stage. When publishing data there is a legal aspect that needs to be addressed and hence the relation here to legal services. In addition to this, the use and reuse of data is tightly coupled to analysis and computational services.
Transform	Computation, Analysis and Visualization; Preservation	Transforming the data relates to producing new derived version of them by either analysis, visualization or for preservation purposes.

Table 2. Mapping the data management and curation services framework to the DCC Curation Lifecycle

5. Discussion of results

The consultation with service providers by means of the interviews and the workshop has helped to understand where expertise is to be found, the services on offer as well as to define in more detail the services needed by researchers at Oxford to create, manage, use and share their data. Although the focus of this document was on services required by researchers, several service units have expressed their need to have better guidance on how to best provide those services. Below the provision of services in Oxford is looked at within each of the different research data management services.

Data management and sharing plans

Currently there is no coordinated service provision for researchers to get advice on how to produce the data management and data sharing plans. As seen on the scoping study interviews, researchers tend not to plan their data management at the outset of the project but they will increasingly be asked to do so. Researchers need to be advised at the funding application stage.

Legal and ethical

Different service units provide some level of legal advice for data creation, sharing and use but those services don't tend to be made explicit. When more complex legal and ethical issues arise service units tend to refer to Legal Services and the Central University Research Ethics Committee (CUREC). Legal Services offers help on IP related issues whilst CUREC provides guidance on the ethical review of research projects involving human participants. OUCS, through the IT Learning Programme, also provides courses on copyright in online resources for content authors and users. Therefore, there are many services and pockets of expertise but a more joint up approach could benefit researchers.

Best formats and best practice

There is some advice provided on best practice and best formats to create some forms of research data through the Research Technologies Service. Other service units also offer advice on best practice but, again, without making the services explicit. OUCS, through the IT Learning Programme provides courses on database design, structure and querying. In sum, although there are some services and expertise, advice on best format and best practice would need to be extended, coordinated and linked to digital preservation services.

Secure storage

In terms of secure storage, it seems like local units such as departments and colleges provide the storage while OUCS provides the back up service and the long term archival. The OSC storage resources are also being used because those researchers producing vast amounts of data have nowhere to put them and researchers also use the storage available in NGS. There is a real and pressing need for providing a more comprehensive secure storage service that links to other services such as OSC computing resources, the Microsoft SharePoint groupware solution and others.

Metadata

OULS units collect metadata about their digital collection and they have great expertise on the range of metadata formats and functions. The RTS also provides some advice on metadata for some types of datasets. Nonetheless there is not a single place where researchers can go to get advice on different standards available and there are no tools for them to create and store their metadata for their datasets.

Access and discovery

Services such as the Social Science Data Library or the Social Science Library provide access to a wide range of data resources and they also assist with data discovery. Nonetheless this type of support and tools are not available in other disciplines.

Computation, Analysis and Visualization

The OeRC provides computational power and analysis tools through OSC, NGS and OxGrid. In addition to this, a visualization service is planned to be provided in the near future. OUCS and others across the University provide training courses for using some of the tools to analyze and manipulate data such as SPSS, STATA, MapInfo, Matlab or NVivo.

Restricted sharing

IMSU helps researchers from the Medical Sciences Division by setting up FTP servers to allow them to share data and the Computing Laboratory is developing some middleware through the GIMMI project that is being tested in medical databases and that has the potential to be made available more widely. NGS also allows its users to share data in a controlled way using grid security mechanisms. Nonetheless, there is no solution that allows sharing large datasets with fine-grained access controls for most divisions.

Data cleaning

Although the Social Science Data Library offers data cleaning services, this type of service is again not available to researchers from other disciplines so that they can prepare data to be sent to domain archives or get advice on what data anonymization practices they can use to protect the confidentiality of their research subjects.

Publication

The ORA has started publishing *small datasets* and metadata of big datasets held at the Oxford e-Research Centre and ODL publishes their data collections on the web. Nonetheless, researchers keep publishing their data on departmental websites or not publishing their data at all. In some cases this is due to the lack on appropriate infrastructure in Oxford to securely and reliably publish all the types of data that researchers are producing.

Assess value

Currently departments or researchers keep data and they assess the value of these resources but there aren't any central services apart from the guidance that RSO may provide for retention periods of research outputs required by the different funding bodies.

Preservation

Some of the OULS units like ORA and ODL are involved in digital preservation activities to maintain access to their digital collections over time. Digital preservation services need to be provided in close collaboration with other services like advice on data management and sharing plans as well best format and best practices.

Add value

There are a few examples on adding value to research data like OULS work with their digital collections and the tools that BVREH project is developing for image annotation.

Each of the services in the proposed framework can now be organized depending of the level of service provided in Oxford into low, medium and high. Low level of service contains the services that are being provided in an ad-hoc basis, they are not being made explicit or they are not offered at all. Medium level of service includes services which are made explicit but that in many cases do not cover all divisions or the capabilities required by researchers. Those services at the high level of service include advance services being provided which would benefit from coordination with the other services in the framework.

As shown in the figure below, most of the services are offered in an incomplete basis, i.e. not covering all disciplines, with the complete functionality required or not offered at all. Those that are being offered at the highest level need to be coordinated with the others to make those services more seamless to researchers using them, improve discoverability and promotion of services.

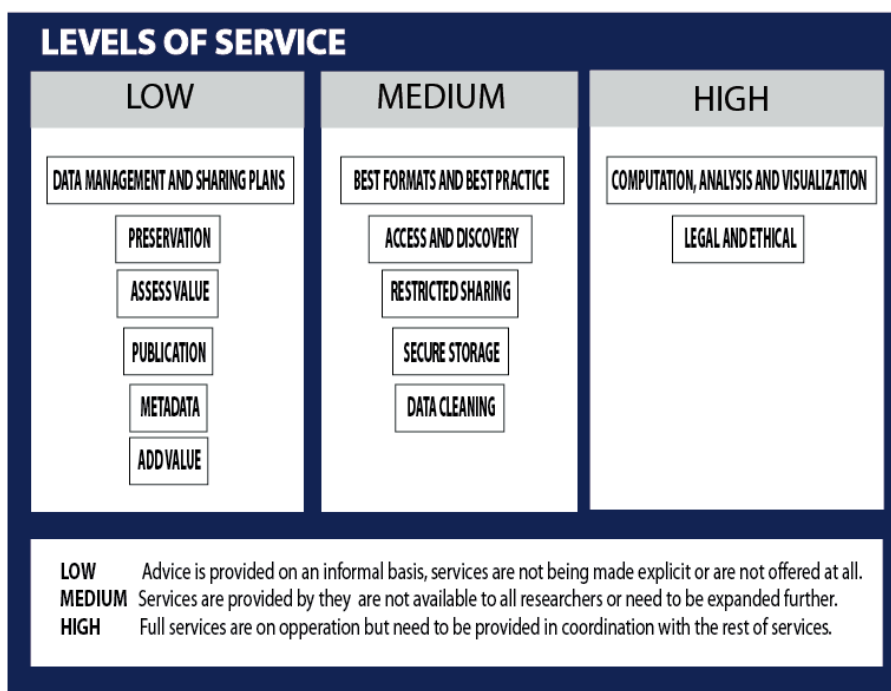


Figure 3. Levels of service for the research data management services

From the above findings it is possible to conclude that:

Expertise in the different areas of data management and curation is widespread amongst central service units at Oxford. In many cases support is provided in an individual and ad-hoc basis but services are not being made explicit as per the lack of resources to offer those.

On the whole, the vast majority of the research data management services identified are not being offered fully or at all by service units across the University. Moreover, those services currently on offer would benefit of coordination to make them more seamless to the researchers using them.

In addition to this, there is a need for a University wide policy for data creation and management as well as for provision of advice and guidance to service units at all levels.

6. Next priorities

Another activity that is currently taken place as part of the Scoping Digital Repository Services for Research Data Management project is the piloting of the Data Audit Framework Methodology in Oxford. Funded as part of the JISC DISC-UK DataShare project, this exercise aims to learn about data management practices within research groups, to identify data assets and to pilot the methodology in Oxford. Two research groups are participating in the exercise: the Young Lives project from the Department of International Development and the Cardiac Mechano-Electric Feedback Group within the Department of Physiology Anatomy and Genetics.

The UK Research Data Service (UKRDS) feasibility study is presenting its final report in its international conference in February. The UKRDS is proposing a Pathfinder phase with a small group of key stakeholders, including the University of Oxford as one of the original case study sites and Oxford is committed to playing a significant role within the foreseen UK Research Data Service Pathfinder activities

7. Recommendations

The findings from the scoping study interviews, the consultation with service providers and the two workshops were presented to the Oxford Digital Repositories Steering Group (ODSRG) in the November meeting. The discussion produced the following two main recommendations and ODRSG members felt that these needed to have clear business cases:

I. One FTE post should be created to be proactive in raising awareness and provide support for research data management.

The post should be based at OUCS, OeRC or OULS and should be funded from central funds. The person in post should have a thorough understanding of the requirements of research data funding bodies. They should also be responsible for raising awareness and training research facilitators to produce data management plans and obtain funding for data management activities.

II. Infrastructure services should be developed to support the research data lifecycle recognizing that resources already exist.

JISC funding would be sought to deploy a prototype which will integrate the necessary existing infrastructure to solve the research data management needs of a research community in Oxford (the Computational Biology Group led by Prof. David Gavaghan). The experience from this prototype will feed into the development of more general infrastructure services to meet the needs of researchers across disciplines.

APPENDIX 1. Framework for Consultation with Service Providers

Introduction

Give brief introduction to the Scoping Digital Repositories Services for Research Data Management including overall aim and objectives. Provide an overview of the questions that will follow and remember the interviewee about the nature of the semi-structured interview, the intention of taking notes, record the interview (with permission) and to publish findings.

Interview

1. Could you explain briefly what is your role within your department?
2. Could you describe the services your department provides for researchers focusing on those that are related to data management?
3. Could you provide examples of those services?
4. The top requirements from Oxford researchers for services to help with their data management activities gathered from the interviews and the workshop are:

Advice on practical issues related to managing data across their life cycle. This help would range from assistance in producing a data management/sharing plan; advice on best formats for data creation and options for storing and sharing data securely; to guidance on publishing and preserving these research data.

A secure and user-friendly solution that allows storage of large volume of data and sharing of these in a controlled fashion way allowing fine grain access control mechanisms.

A sustainable infrastructure that allows publication and long-term preservation of research data for those disciplines not currently served by domain specific services such as the UK Data Archive, NERC Data Centres, European Bioinformatics Institute and others.

Funding that could help address some of the departmental challenges to manage the research data that are being produced.

- i. How do these requirements fit with your experience with your users?

Using the Research Data Management Services Diagram:

- ii. How do the existing services currently provide from your department map to those requirements?
- iii. How do your planned services in the next 3 to 5 years map to those requirements?
- iv. Which service providers would you expect to deliver those services?

5. Is there anything else that you would like to add?

De-Brief





6. How do you think the interview went?
7. What are the benefits you believe you get from participating?
8. Could you suggest anyone you know that could participate in these interviews?

APPENDIX 2 – Research Data Management Services Diagram

This diagram is aimed to provide a visual tool that can be shown to service providers to assist with the finding of current and planned research data management services by division.

The top row shows a comprehensive framework of research data management services; these include not only the researchers requirements for services gathered previously but some other services coming from the DCC Curation Lifecycle Model. It is intended that all of those services cover a situation where the researcher is supported to work with data throughout his research lifecycle but also once the data come under the responsibility of professional data curators. From the fourteen services listed a subset could be categorized as core and the others as optional. The subcategories *tools* and *advice + support* are used to clarify each of the services and more could be added if needed.

The research data management diagram will be shown to service providers to see which services they provide and for what divisions. National and international facilities such as UKDA, the NERC Data Centres or EBI can also be added. At the end of the consultation exercise, the diagram will help visualize what services are, or will be provided and where the gaps are.

RESEARCH DATA MANAGEMENT SERVICES	DATA MANAGEMENT/ SHARING PLANS	BEST FORMATS AND BEST PRACTICE	STORAGE AND BACK UP	ANNOTATION	ACCESS	ANALYSIS	VISUALIZING	RESTRICTED SHARING	LEGAL	PREPARING DATA FOR PUBLICATION or DEPOSITION	PUBLICATION	ASSESSING VALUE	PRESERVE	ADD VALUE
			TOOLS	TOOLS	TOOLS	TOOLS	TOOLS	TOOLS		TOOLS	TOOLS	TOOLS	TOOLS	TOOLS
OXFORD DIVISIONS	ADVICE + SUPPORT	ADVICE + SUPPORT	ADVICE + SUPPORT	ADVICE + SUPPORT	ADVICE + SUPPORT	ADVICE + SUPPORT	ADVICE + SUPPORT	ADVICE + SUPPORT	ADVICE + SUPPORT	ADVICE + SUPPORT	ADVICE + SUPPORT	ADVICE + SUPPORT	ADVICE + SUPPORT	ADVICE + SUPPORT
Mathematical, Physical, & Life Sciences → 														
Humanities → 														
Medical Sciences → 														
Social Sciences → 														

APPENDIX 3 – Workshop Institutional and National Research Data Management Services

Abstract

This workshop is the second event organised under the umbrella of the Scoping Digital Repository Services for Research Data Management project. The overall aim of the workshop is to hear about examples of support services for research data management in relation to the requirements for services gathered through the interviews with researchers. The event is designed for Oxford researchers to learn about different types of data management services that could potentially be delivered in Oxford as well as for Oxford service providers to discuss those services and their role to support researchers with their data management.

Objectives

- Raise awareness of best practice in institutional and national support data services
- Complement the findings of the scoping study consultation with Oxford support services
- Help to identify gaps in the provision of services available to researchers to assist with their data management duties
- Stimulate discussion amongst service providers about collaborative ways to address data management issues in Oxford

Programme *	
10.00 – 10.30	Registration
10.30 – 10.45	Welcome and introduction Prof. Paul Jeffreys, University of Oxford Director of IT
10.45 – 11.45	<i>Dr. Natasha Balac, San Diego Super Computer Center</i>
11.45 – 12.30	The UK Research Data Service (UKRDS) Feasibility Study - Neil Beagrie, Charles Beagrie Ltd
12.30 – 13.30	Lunch
13.30 – 14.00	Tools for managing research data: the Data Audit Framework (DAF) and DRAMBORA - Sarah Jones, Digital Curation Centre.
14.00 – 14.30	Data Management Planning: what do researchers really need to know? - Louise Corti, UK Data Archive.
14.30 – 15.00	The Archaeology Data Service, a subject specific digital repository – Dr. Stuart Jeffrey, Archaeology Data Service.
15.00 – 15.20	Beverages
15.20 – 15.50	Building tools to handle environmental "omic" data - Tim Booth, NERC Environmental Bioinformatics Centre.
15.50 – 16.20	Legal issues relevant to digital repositories – Carolyn McKee, University of Oxford Legal Services
16.20 - 17.00	Service Units Roles Panel Discussion - Chaired by Prof. Paul Jeffreys. Dr. Mike Fraser (OUCS), Sally Rumsey (OULS), Kathryn Dally (RSO), Neil Beagrie (UKRDS) , Dr. Jon Lockley (OeRC)

*Slides can be found at: <http://www.ict.ox.ac.uk/odit/projects/digitalrepository/Workshops-DataServices.xml>