



LIBER Case Study:

University of Oxford Research Data Management Infrastructure

AUTHORS: Dr James A. J. Wilson, University of Oxford, james.wilson@it.ox.ac.uk

KEYWORDS: generic, institutional, software services, support, training, infrastructure, JISC MRD

1 What was the starting point?

Work on an institutional approach to research data management infrastructure for the University of Oxford began in around 2009 as several groups within the university began developing tools and processes to help store and document data, sometimes with JISC funding. By 2011 it was becoming apparent that the groups involved in the development of 'generic' (as opposed to discipline-specific) RDM services would benefit from fully collaborative working and improved coordination of activities to ensure that the aspects of infrastructure that each group was working on could pass data and metadata to related services intended for researchers at different stages of the research data lifecycle. The principle groups involved in the development of institutional infrastructure were the Bodleian Libraries, IT Services, Research Services, and researchers from the Department of Zoology. With funding from the JISC Managing Research Data Programme, a collaborative project entitled 'Data Management Roll-Out at Oxford' (Damaro) was launched. Although the project concluded in June 2013, work on developing the integrated RDM infrastructure at the University of Oxford continues.

The drivers behind the RDM infrastructure projects have been the major research funders (particularly the EPSRC) and, to some extent, researchers themselves, many of whom recognise the importance of sound data management but do not feel that they are necessarily *au fait* with 'best practice'. Interviews and surveys suggest that researchers would like support with a number of different aspects of data management, including planning, organising, security, sharing, preparing datasets for deposit and long-term preservation, and issues relating to copyright, licensing and intellectual property.

2 What kind of research data is targeted?

Whilst teams within the Oxford eResearch Centre and several of the academic departments are involved in developing discipline-specific research data management tools, at the institutional level the emphasis is rather on assisting a broad range of disciplines by developing relatively generic tools, most of which cater less for 'Big Data' than the long tail of data outputs which tend to be less well-supported at present.

3 What is the organisational framework?

Roles and Responsibilities

The University of Oxford is working on a number of component parts for its infrastructure. These include

- Oxford DMPOnline: Online templates to assist researchers in completing data management plans.



June 2014

- DataStage: Open source software enabling research groups to set up shared drives with different levels of access for collaborators. DataStage facilitates the deposit of data (with accompanying metadata) into ORA Data and other data repositories.
- The Online Research Database Service (ORDS): An in-development cloud-hosted service enabling researchers to collaboratively create, edit, and share relational databases. Subsets of databases can be published and cited. Again, the system facilitates the deposit of data (with metadata) in ORA Data and other data repositories.
- ORA Data: ORA Data (Oxford University Research Archive Data) consists of two elements, a data archive and a data finding tool. The data archive provides a long-term home for research data created by researchers at Oxford for which more appropriate subject-specific repositories are not available. DOIs will be assigned to data deposited in the Archive. The data finder tool consists of a catalogue of research data produced at Oxford, whether hosted in ORA Data or elsewhere. It can harvest metadata from external sources and includes a data-deposit interface where researchers can enter or edit metadata manually.

In all of the above cases the software is either mature or close to reaching the point at which it can be launched as a service. Some of the services, e.g. ORA Data are already running in a limited mode. Service staffing and support are not yet secure. Besides centrally-provided services, researchers already use a wide range of externally provided software and services to manage their data, and many disciplines have national or international data repositories where they can deposit data. The University is looking to broaden awareness of these and integrate systems where possible.

Besides developing software, the University has also been developing support and advisory services. These include a central RDM website for the University, intended as a first port of call for researchers wishing to learn more about research data management (<http://researchdata.ox.ac.uk>), and a new cross-departmental advisory service staffed by members of IT Services, Research Services, Bodleian Libraries, and the Oxford e-Research Centre.

It became apparent fairly early in our information-gathering exercises that expertise in the various aspects of research data management is distributed between the main support departments (Research Services, IT Services, and the Bodleian Libraries) and the academic departments and faculties. Few, if any, staff have a great deal of confidence offering advice relating to all phases of the research data lifecycle, although many individuals have some degree of expertise in particular aspects of research data management.

A researcher survey that we ran indicated that researchers have very varied expectations regarding who they should turn to for advice relating to research data management. Asked where they would go to find information about research data management, the most common responses were IT Services, departmental staff, and fellow researchers, but even IT Services was specified by fewer than a quarter of those who hazarded a response.

Oxford is therefore taking the approach that the institution should offer a clear and well-publicised single point of contact for researchers with queries relating to research data management, but those tasked with monitoring the email address will then pass on any queries that they cannot straightforwardly address themselves to individuals with the appropriate expertise in whichever department they may be based. The single point of contact enables the institution to track common questions and prepare FAQs and other materials for the University's RDM web pages. A 'directory of expertise' has been set up to record who within the University can help with different types of enquiry, and a mailing list is used to discuss new or unusual issues.

Besides advice, Oxford is also developing software services to support RDM. The model we are considering regarding supporting and maintaining these is that services relating to the planning

and bidding phases of the research data lifecycle should be overseen by Research Services, services primarily relating to research data management during the ‘live’ phase of a project (such as the ORDS) should be managed by IT Services, and services relating to the post-project data curation, preservation, discovery and access (e.g. ORA Data) should be managed by the Library Services, who have the most experience in dealing with this sort of information.

The University is still considering how to address the long-term financial sustainability of the services it is implementing. It is likely that each service will have its own resourcing arrangements, so that those services that have the potential to pay their own way can do so whilst central funding is reserved for obligatory general services that have little possibility of recouping their costs by other means. This modular approach will hopefully also enable new services to be added or existing services removed in response to demand, without disrupting other aspects of the infrastructure.

Policies

The University of Oxford Policy on the Management of Research Data and Records was ratified in July 2012. It was developed by Research Services in collaboration with the other support departments and is ‘owned’ by the University’s Research and Information Sub-Committee, a sub-committee of the University Research Committee. It is a high-level policy that explains the rationale for research data management and defines the responsibilities of researchers and of the University in supporting researchers. The policy refers to the University’s ‘obligations under research funders’ data-related policy statements and codes of practice to ensure that sound systems are in place to promote best practice, including through clear policy, guidance, supervision, training and support. It also recognises, however, that the motivation to manage research data well is not merely derived from the need to conform to funding mandates, but additionally that ‘research data are valuable to researchers for the duration of their research, and may well have long-term value for research, teaching and for wider exploitation for the public good, by individuals, government, business and other organisations.’

The Policy specifies that research data and records should be:

- a. Accurate, complete, authentic and reliable;
- b. Identifiable, retrievable, and available when needed;
- c. Secure and safe;
- d. Kept in a manner that is compliant with legal obligations and, where applicable, the requirements of funding bodies and project-specific protocols approved under the University Policy on the Ethical Conduct of Research Involving Human Participants and Personal Data
- e. Able to be made available to others in line with appropriate ethical, data sharing and open access principles.’

The Policy states that the minimum retention period for research data and records is ‘three (3) years after publication or public release of the work of the research’, but also notes that in many cases researchers will want to preserve their data for longer, and that if funder policies require data to be kept for longer, it is their requirements that have precedence.

The University Policy places the responsibility on researchers to develop and document clear data management procedures for their research, including enabling post-project curation. The University, meanwhile, is responsible for providing access to ‘services and facilities for the storage, backup, deposit and retention of research data and records’, providing ‘training, support and advice’, and providing ‘the necessary resources to those operational units charged with the provision of these services, facilities and training.’

The Policy is available via Oxford's central Research Data Management Website, at <http://researchdata.ox.ac.uk/university-of-oxford-policy-on-the-management-of-research-data-and-records>.

4 What kind of support services are provided to researchers?

The University already offers researchers advice and support with several aspects of research data management [see above] and is working to improve the skills and expertise of support staff as well as researchers.

A number of interviews with and surveys of researchers have been conducted over the last few years. The largest survey was the University-wide Research Data Management Survey conducted in November and December 2012. Smaller surveys have sought to understand the RDM training needs of scientists, current practices and requirements in the Humanities Division, the confidence of support staff when dealing with RDM questions from researchers, departmental practices and challenges, and requirements-gathering surveys and interviews for individual services.

Several of the RDM projects at Oxford have worked with specific researchers and research groups, particularly in relation to the development of software tools. Projects have included collaborations with heart scientists generating large three-dimensional images of hearts (and derived data), with classicists studying sites of economic activity in the ancient world, with researchers analysing 20th century human migration data, and a range of others. That said, much of the development that has taken place has focused on 'generic' solutions that can be applied across disciplines. The involvement of particular research groups has, until recently, tended to be largely focused around requirements gathering and feature prioritisation, testing, and validation.

Thus far most development of research data management expertise amongst staff has come about due to their involvement in the projects we have run (which have been a learning experience for us all) rather than through specific training programmes, although we are now starting to consider how best to train subject librarians.

5 What kind of infrastructure is provided?

As previously indicated, the focus of the Oxford RDM projects has been on developing software services along with guidance and training for researchers. Efforts have been directed across the full research data lifecycle. In most cases the software is either mature or close to reaching the point at which it can be launched as a service. The initial set of services for RDM should cover data management planning, relational database support for researchers during the live and post-project stages of their research, assistance with documentation, data registration and preservation, and data sharing and publication (to enable re-use). A particular focus has been on ensuring that data and metadata can be passed from one component to another to improve metadata quality and minimise the amount of manual input required of researchers themselves. The software services have been developed in-house, but that does not mean that the university will not consider supplementing them, or even replacing them, with externally-developed or commercial solutions in future if such solutions are deemed preferable.

Most of the services being developed are intended for principal investigators based at the University of Oxford and their collaborators (regardless of where their collaborators are based). Data held in the systems under development can be shared between a specific group of individuals or opened to the general public if required or so desired by the researchers responsible for the data.



6 What have you learned so far? What's next?

Due to the nature of large academic institutions such as the University of Oxford, there is necessarily a broad range of 'stakeholders' in research data management. It is an issue that affects anyone undertaking research or making decisions about research administration. As such stakeholders include most senior academic staff and most support departments. With so many stakeholders involved, each bringing their own particular interests and understanding of the issue involved to bear on the topic, it can be a challenge to ensure buy-in. In particular, researchers tend to have different views on the requirements or necessity of any given component of RDM infrastructure depending on their disciplinary background. Those from disciplines such as astronomy, physics, crystallography and so forth, where researchers have generally already recognised the benefits of data sharing and put in place an appropriate infrastructure, may be uncertain of any benefits that central support can bring. Others, particularly those from disciplines where data is very heterogeneous or where researchers often manage their data as individuals, may regard institutional support with RDM as welcome, given that many are uncertain as to what constitutes good practice.

Much of the work of setting up an RDM infrastructure therefore involves gaining a clear understanding of what is required by whom and ensuring that those who will be involved in resourcing and supporting the infrastructure are aware of the rationale for implementing it.

The difficulty of creating a quantitative business case for such an infrastructure is also a challenge. Whilst the costs are generally fairly straightforward to quantify, many of the benefits are intangible – relating to risk reduction, improved practices, better institutional awareness of assets, increasing the likeliness that over the medium or long term data is re-used, and so forth.

Oxford has established a Working Group with reporting lines within the existing committee structure to steer development and ensure senior management remain informed, as well as to ensure the involvement of all academic divisions. The Working Group has so far helped to prioritize which requirement should be met and advised on business models.

The next challenge that the University needs to address is resourcing the ongoing services required to meet pressing funder requirements. There is also more work to be done to ensure that the coordination of RDM support services and guidance is able to respond to the needs of researchers in an optimal manner, drawing on expertise embedded in multiple support departments, but also ensuring that gaps in knowledge are identified, and spreading awareness of good practice more generally. A recent push to include data security expertise in the support group should enable us to address researchers' concerns in that regard more effectively, and we continue to add content to the website.

Further information

WEBSITE: <http://researchdata.ox.ac.uk/>

CONTACT: Dr James A. J. Wilson, University of Oxford, james.wilson@it.ox.ac.uk