

The International Journal of Digital Curation

Volume 8, Issue 2 | 2013

Implementing the Research Data Management Policy: University of Edinburgh Roadmap

Robin Rice, Cuna Ekmekcioglu, Jeff Haywood, Sarah Jones, Stuart Lewis,
Stuart Macdonald and Tony Weir,

Information Services,
University of Edinburgh

Abstract

This paper discusses work to implement the University of Edinburgh Research Data Management (RDM) policy by developing the services needed to support researchers and fulfil obligations within a changing national and international setting. This is framed by an evolving Research Data Management Roadmap and includes a governance model that ensures cooperation amongst Information Services (IS) managers and oversight by an academic-led steering group. IS has taken requirements from research groups and IT professionals, and at the request of the steering group has conducted pilot work involving volunteer research units within the three colleges to develop functionality and presentation for the key services. The first pilots cover three key services: the data store, a customisation of the Digital Curation Centre's DMPonline tool, and the data repository. The paper will report on the plans, achievements and challenges encountered while we attempt to bring the University of Edinburgh RDM Roadmap to fruition.





Introduction

University obligations in the area of research data management are explored in a previous paper, which conveys the direction at the University of Edinburgh (Rice & Haywood, 2011). This has led to the creation of the first institutional research data management policy by a UK Higher Education Institution (HEI) in May, 2011.¹

The paper discussed a number of drivers for the university's focus on research data management and storage, including adoption of the UK Research Integrity Office's Code of Practice for Research (UKRIO, 2009) and the need for principal investigators (PIs) to comply with funders' requirements. Since that time another UK funder, the Engineering and Physical Sciences Research Council (EPSRC), has issued a Policy Framework on Research Data (EPSRC, 2011) with significant requirements addressed to the institution, not just the PI. This has raised the bar for Edinburgh and other universities, and encouraged research institutions to be focused in their commitment to Research Data Management (RDM) policy and infrastructure.

The University of Edinburgh, one of four ancient Scottish universities and a leading European centre of academic excellence, is a research-led institution with a mission for the 'creation, dissemination and curation of knowledge'. Information Services (IS), a support group within the university, provides support for the research and education activities of schools across three colleges, including a Data Library service, and provides two UK national services (EDINA and Digital Curation Centre) as well as hosting the Scottish Digital Library Consortium.

The Vice Principal Knowledge Management and Chief Information Officer, Jeff Haywood, has declared research data management and storage to be a major IS-led project by committing significant capital funds to this effort. An IS cross-divisional RDM Policy Implementation Committee has been formed, charged with delivering those policy objectives which can be achieved over an 18-month period beginning August 2012. An RDM steering group, made up of academic and support group representatives, chaired by Professor Peter Clarke from the School of Physics, has been established to provide oversight of the activity of the Implementation Committee, ensuring that the services meet the needs of researchers in every school (the sub-division under the three colleges), through pilot and policy initiatives. Concurrently, the Digital Curation Centre (DCC) is running a series of institutional engagements to support universities to build their RDM capacity and capability.² The University of Edinburgh is one of 20 institutions taking part in this programme.

University of Edinburgh RDM Roadmap

The implementation strategy is set out in the Research Data Management Roadmap, a dynamic document with indicative actions and deliverables over the first 18 month period (University of Edinburgh, 2012).³ This paper discusses ongoing and planned

¹ University of Edinburgh research data policy: <http://www.ed.ac.uk/is/research-data-policy>

² DCC institutional engagements: <http://www.dcc.ac.uk/community/institutional-engagements>

³ Due to changes in university planning procedures, funding was delayed until July 2013. The Roadmap has been updated accordingly.

work under the four main roadmap headings: Data Management Planning, Active Data Infrastructure, Data Stewardship, and Data Management Support.⁴

The RDM work at the University of Edinburgh continues to be informed by a wealth of national and international initiatives and activities, none more so than the JISC Managing Research Data (MRD) Programme (2011-2013)⁵ and its predecessor programme, aimed at producing software solutions, supporting systems, guidance and policies to benefit universities (Beagrie, 2011).

The recent Royal Society report ‘Science as an open enterprise’ (2012) states as one of its recommendations:

‘Universities and research institutes should play a major role in supporting an open data culture by ... developing a data strategy and their own capacity to curate their own knowledge resources and support the data needs of researchers.’

The Association of European Research Libraries (LIBER) also recently published ‘Ten recommendations for libraries to get started with research data management’ (2012). Both reports confirm the university’s policy principles and provide further guidance for IS to align its RDM work with their recommendations.

Data Management Planning

The main focus of DCC support is in leading the Data Management Planning pilot work. Research funders increasingly expect researchers to submit data management and sharing plans (DMPs) in their grant applications. The University of Edinburgh roadmap has two objectives to help researchers meet these requirements.

1. Tailored DMP assistance for PIs submitting research proposals

We are currently testing desire for consultancy services to assist researchers with writing data management plans (DMPs). The need for such a service has been predicted to be strongest in the College of Humanities and Social Sciences, where queries about whether the research outputs constitute ‘research data’ and whether research proposals require a DMP are commonplace.

To date, we have been consulting with Edinburgh Research and Innovation (ERI) to raise awareness of existing services and ensure researchers are directed to them during the grant proposal process. Researchers often ask for example DMPs and a number of successful proposals have been shared with us so we can draw out good practice. We will provide a library of successful DMPs gathered from ERI and the schools, and explore options to expose these examples via an Edinburgh version of the DCC’s DMPonline tool.

⁴ Progress on the Roadmap can now be tracked through the Edinburgh Research Data Blog: <http://datablog.is.ed.ac.uk/>

⁵ Jisc Managing Research Data programme: http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata.aspx



2. Customise DMPonline for optimal University of Edinburgh use

The focus of the DMP pilot at Edinburgh is on the use of DMPonline. We began work by contacting existing users and asking for their feedback on the tool. Around half of the 24 users registered with a University of Edinburgh alias were research active and half were in research support roles. We asked them why they had registered for the tool, what was their experience of using it, and whether they would like to be provided with examples or more tailored advice. We have since followed up by running a focus group and holding a number of guided interviews with researchers. To balance any bias in the feedback, we also held usability tests with researchers who had not encountered the tool before.

Most users had registered out of curiosity to see what the tool does and whether it could be of use to them. Unfortunately, many researchers were confused by the number of options presented or frustrated by the level of detail. One researcher explained that he expected a more straightforward tool, with pre-populated answers to explain the kind of generic information that needed to be included along with project-specific issues. Some researchers resorted to contacting ERI for example plans or asking colleagues and support staff for advice instead.

Despite these concerns, there was positive feedback about the workflow and how the tool has improved over time. People were in favour of having an online tool and could see lots of opportunities to enhance DMPonline with details of local support and practical examples to help researchers interpret the questions. The tool was also felt to be well-coded; the concerns noted reflect flaws in the conceptual framework.

The DCC has reviewed the feedback internally and is planning for redevelopment.⁶ There is enough buy-in at Edinburgh or researchers willing to work with us to revise the tool. We expect to regularly test amendments to ensure they are fit-for-purpose and meet researchers' needs. We also plan to roll out the tool school by school from autumn 2013, working with each group to identify local support and good practice to develop school-level guidance for incorporation into DMPonline as far as possible.

Active Data Infrastructure


3. To provide a globally accessible cross platform file store with sufficient capacity to satisfy majority of researcher use cases

From consultation with the university's research community, the provision of free-at-point-of-use storage to researchers was identified as a key service deliverable. This would provide the working space for data for the majority of university research, with the option to extend allocation through direct payment.

A free-at-point-of-use service removes one of the main barriers to good stewardship and allows data to be held on the appropriate class of infrastructure, rather than one selected on lowest cost alone. This provides the assurance and security of a centralised solution with full resilience and multi-site recovery, but also leverages the institutional savings of large scale procurements and centralised staffing.

The university has provided a centralised research file store as a pay-for-use service in recent years. This has allowed a large degree of expertise to be developed in

⁶ The DCC announced its plans in a blog post in January 2013 (Ashley, [2013](#)).



the technologies required to deliver petabyte-scale storage infrastructure, and also provided the detail on user requirements.

There is great variety in expected use of the active data store, such as general document handling; a holding area for very large data sets, which will be staged onto High Performance Computing systems for analysis; storage for instrument data such as MRI scanners or gene sequencing systems.

Developing an infrastructure to support this variety of uses has been challenging, and has led the university to implement an open standards-based storage infrastructure which is scalable and able to utilise storage from many hardware vendors, rather than an appliance solution, which would be tied to delivery from one vendor. This flexible delivery supports layering of a range of data access mechanisms on top of a common file store.

The costs involved in such institutional deployments are substantial, and close regard must be paid to the sizing of the service to ensure it meets needs. A previous university research computing survey (Ekmekcioglu, [2007](#)) provided valuable information on the size of working data sets in different domains, and this was scaled to provide expected service dimensions. At the time of writing, a full design is in place with detailed service risk register and service level definition approved by the RDM steering group. Initially, the access mechanisms appropriate to file delivery to common desktop environments will be deployed, i.e. CIFS and NFSv3. Additionally, sshfs (SSH Filesystem) will be used to deliver access to files outside the university's network environment.


There are several ongoing challenges to work through. The allocation model needs to ensure that stakeholders achieve equitable value whilst ensuring those areas with greater data requirements are satisfied. Another challenge is in anticipating changing cultural attitudes to data management, with responsibility moving from the individual to a defined role of 'data manager' within the research group.

4. Provide additional data access mechanisms to better support mobile devices and external collaboration

Researchers do not work solely on fixed desktops sited on the university's network. Much research is collaborative and requires access from researchers at other institutions. After the initial delivery of the active file-store, the service will be developed to better support mobile clients and external collaborators.

'Bring your own device' access to central services is a common theme on campuses, with service access moving from fixed desktop, to laptop to mobile device. Whilst it is unlikely that very large datasets will be required to be delivered to handheld devices, it is now commonly expected that data will be available on personal smart phones or tablets. A range of access mechanisms will be considered to mature the service, including WebDAV.

'Dropbox-like' access to data was a commonly stated requirement in the consultation. Developing a new service where a free internet facility is already available must be carefully considered. A path-finder project to understand the need for such a service found that researchers used such technologies for simple network



data access, for synchronisation across a number of personal devices and for collaborative access to shared data.

Currently, external collaborator access to data requires visitor registration with the university's identity management system. To provide greater flexibility, federated access via Project Moonshot⁷ will be investigated.

5. To provide mechanisms to address backup and synchronisation of mobile devices

As the university continues to improve the central hosting of data, the need to secure data held on remote devices may lessen. However, the need to improve the recovery of data held on mobile devices remains, either through data synchronisation techniques or via data backup.

6. Provide a service to ensure integrity and long term retention of golden copy research data

This objective sits between the dynamic active data storage system and the data sharing repository: the requirement to provide long term storage of 'golden copy' research data. That is, data that needs to be stored in a final state for a long period. The important aspects of this kind of service are long term storage, rights metadata, security, and preservation.

The data need to be stored for a long period. Some datasets may have retention schedules (delete after x years), others may need to be kept indefinitely, whilst others will need to be kept for as long as they are being accessed. Rights metadata is required to ensure that data held within the vault can be identified, maintained and re-used as appropriate. Ensuring metadata is kept up to date will be a challenge in the long term, as data ownership changes with staff turnover. Some data will be stored in a data vault rather than an open research data repository due to the sensitivity of the data. It is therefore essential that such a system can enforce high levels of security in authorising access to these data. Preservation in this context covers the refresh of hardware systems as old storage systems are retired and replaced with new systems, and the routine monitoring of files to ensure that they are in the same state as when originally deposited. Continuing access is a roadmap objective in itself (see point 10).

Data Stewardship

7. To develop the data repository for enhanced deposit and discovery of data collections generated by university researchers

Edinburgh DataShare⁸ is an institutional data repository set up by the Data Library to allow University of Edinburgh researchers to deposit, share and license their data resources for online discovery and use by others, and as such is a key component of data stewardship.

In the summer of 2012, three research groups from Clinical Psychology, Linguistics and English Language (both from the School of Health and Social Sciences) and the Centre for Speech Technology Research (an interdisciplinary

⁷ Project Moonshot: <https://community.ja.net/groups/moonshot>

⁸ Edinburgh DataShare: <http://datashare.is.ed.ac.uk>

research centre) agreed to be pilots to determine how the data repository, built on a DSpace platform, will meet the needs of Edinburgh researchers as depositors. This in turn identified a number of user requirements including guidance on user authorisation and permissions, intuitive documentation and licensing, supported filetypes, batch import and batch metadata production, display of rich media material, including possible streaming. These user requirements are being transformed into a functional requirements specification for implementation early next year to be followed by further testing by current pilots and other new users.

8. To provide a registry of research data assets in support of the university RDM policy

The university policy states that:

‘Any data which is retained elsewhere, for example in an international data service or domain repository, should be registered with the university.’

Emerging international registries of data repositories could be used to discover and recommend appropriate discipline-specific repositories. These include re3data⁹ funded by the German Research Foundation and Databib¹⁰ from Purdue University.

While national developments towards an Australian-style ‘Data Commons’ registry will be watched with interest, the main challenges will be to decide where the metadata records will be held (candidates include the data repository and the CRIS – see below), and how to obtain the information without overdue burden on researchers.

9. To ensure efficient interoperation between all RDM systems as well as PURE

Each new system or process developed to improve the management of research data needs to be integrated into the existing architecture of university systems, such that it maximises the benefits it can deliver whilst reducing the cost and time required to use. Obvious examples of this type of integration are that all new tools should be integrated with the institution’s authentication and authorisation systems, allowing users to log in with their existing credentials, and that central database lookup is enabled to avoid duplication and introduction of error. In addition, flows must be devised that can transfer information amongst the following key systems: the Current Research Information System (CRIS), the data store, the data ‘vault’, and the data repository.

Many institutions run integrated CRIS systems to manage research information. The University of Edinburgh runs Atira Pure, although others exist, such as Symplectic Elements and Avedas Converis. These are typically used to record many facets of the research process, such as funding awards, publications and other research outputs, teaching duties, equipment, and other research-related awards and professional contributions. CRIS systems often act as the database used to deliver official returns to research funders. The recording of records about research data needs to be integrated into the CRIS, either entered directly (like publications or professional research contributions) or automatically through system integrations (like grant

⁹ Re3data: <http://www.re3data.org/>

¹⁰ Databib: <http://databib.org/>

awards fed from the finance systems). Integration with the CRIS allows research data outputs to be managed and related to other aspects of the research process.

A common flow of data will be from a researcher's active data file store into a long terms preservation or access system, such as a data vault or repository. It is therefore important that these are closely integrated to ensure that data can be easily and efficiently transferred between the two. Using a transfer protocol such as SWORD, tools can be written to deposit data directly from the file store into a repository. Alternatively, repositories can describe and provide access to files stored in research data fabric systems, such as iRODS.

10. To provide continuity of access for data assets with long term value

Continuity of access is our preferred term for digital preservation. Central to this objective is the concept of trust, especially the need to keep data safe on behalf of depositors, but also on behalf of current and future users and funders. The data repository administrators will strive for trusted digital repository status through clear policies, procedures and service definitions, and by attempting to achieve a peer-reviewed Data Seal of Approval¹¹ status.

There is a need to work with others in IS, including Special Collections, and to engage in external communities, such as fellow DSpace implementers, to fully realise this objective on behalf of the university. Digital preservation is a moving target because of changing technology, tools and consensus within the field; for example, obsolescence of file formats is no longer considered by some experts to be one of the major problems in long-term digital preservation (Rosenthal, [2010](#)).

The EPSRC Policy Framework on Research Data states that:

‘Research organisations will ensure that EPSRC-funded research data is securely preserved for a minimum of 10-years from the date that any researcher ‘privileged access’ period expires or, if others have accessed the data, from last date on which access to the data was requested by a third party.’ (EPSRC, [2011](#))

This forms part of nine explicit ‘expectations’. Certainly this is one neutral criteria that could be used to assess items with long-term value.


Data Management Support

11. Raise awareness of university and funder policies, and advocate the use of data management plans for all research projects.

A well-planned and engaging awareness raising programme is arguably one of the most effective means of communicating information about ongoing research data management work at the University of Edinburgh, including the university and funder policies.

An awareness raising programme planned by IS will be tailored to three different audiences in schools/research centres/units:

¹¹ Data Seal of Approval Status: <http://datasealofapproval.org/>

- 
1. College management and research committees,
 2. PIs and senior research staff,
 3. Early career researchers and PhD students.

This follows a similar exercise with IS staff during the summer 2012. The first level will deal with high level policy issues. These presentations will be the ‘meta presentation’ to prime the college management and research committees. The second and third levels will involve detailed information about the delivery of specific services (e.g. active data infrastructure, DMPonline, DataShare).

Whether communicated through meetings, workshops, articles or video, the message will be positive, focusing on the opportunities and benefits provided by the new research data management services rather than criticising current practices.

12. Create and revise IS data management guidance

University of Edinburgh was one of the first UK universities to launch online research data management guidance in 2009. This online resource aims to assist university researchers in complying with the increasingly demanding requirements of both external funding bodies and the university, and direct them to appropriate sources of support. The guide has been well received but was considered overly complex. We recently analysed usage of the existing web pages and published a streamlined version¹² that is more directly aligned to the needs of PIs. The new version consists of eight brief pages covering the essential topics researchers must understand before embarking on a research project.

Further work on these pages will involve forming a sub-group to consider costing expectations and sources, and adding research data management costing information for PIs. We are also working with other university colleagues to raise the visibility of these support pages higher up and across the university website, as appropriate.

13. Maintain, develop and promote online training modules

Research Data MANTRA¹³ open online training modules were developed by the Data Library to reflect best practice in research data management grounded in three disciplinary contexts (social science, clinical psychology and geosciences) through needs assessments and user testing. The Institute for Academic Development is helping to disseminate the course to more schools and postgraduate students. Take-up of the MANTRA materials will be monitored in order to allow revision and further development of the modules, such as extending materials to include discipline-specific units. Activity such as face-to-face activities and workshops to improve take-up of MANTRA within existing participating schools is ongoing.

14. Create tailored, on-demand training for research groups and professionals

An extended professional development training programme for liaison librarians at the university is being piloted, covering RDM and how it may be applicable to research practises in the disciplinary areas each represents. This knowledge-transfer exercise includes independent study based on MANTRA, reflective writing, face-to-face sessions with short speaker presentations followed by discussion, and

¹² See: <http://www.ed.ac.uk/is/data-management>

¹³ Research Data MANTRA: <http://datalib.edina.ac.uk/mantra/>

group exercises. The training ‘kit’ will be released with an open licence for other institutions, particularly those without local experts on staff, to utilise for ‘DIY’ RDM training, in spring 2013¹⁴.

Two schools have already approached IS to facilitate an RDM workshop; on-demand training for research groups could be done in partnership once the awareness raising headline-level campaign makes inroads into the schools.

15. Trial an in-depth data management consultancy service

Part of this objective is to imagine what an in-depth service for research data management would look like: could the University of Edinburgh embrace concepts such as embedded librarians as part of research projects to work on metadata, or assist IT staff to act as ‘data scientists’ for periods of time to improve data workflows?

Another aspect involves meeting in-depth requests for assistance that are time consuming, including assisted deposit of large or complex datasets, and evidence gathering for potential RDM contracted services for grant-funded projects. Costing research inevitably arises here as well, to determine both the potential demand and benefits for such in-depth RDM service.

Conclusion

The development of an RDM roadmap, agreed with senior academics across the university and signed off as part of a business case by university senior management, has prepared the way for immediate and recurrent investment into both data storage (mainly capital expenditure) and data management (mainly revenue expenditure, i.e. staff). As we develop this roadmap the ‘devils in the detail’ have become clearer, and led to an incremental approach in which we road test infrastructure and services in subject areas to be sure we devise useful and locally valid solutions rather than impose a one-size-fits-all compromise. As part of our internal planning we have been observing and engaging with external developments, especially in the UK, but at the present time we do not expect those developments to offer solutions quickly enough to preclude the need for major local investment. We hope in the near future to reduce some of our data storage commitments in favour of more effective shared solutions.

These are exciting times for IT staff and librarians as we work together planning to support colleagues in research, enabling compliance with funders’ requirements and, most importantly, building up digital collections today for future research.

References

- Ashley, K. (2013). *Future plans for DMPonline*. DCC Blog. Retrieved from <http://www.dcc.ac.uk/news/future-plans-dmponline>
- Beagrie, N. (2011). *Benefits from the infrastructure projects in the JISC Managing Research Data programme*. Retrieved from <http://www.jisc.ac.uk/whatwedo/programmes/mrd/outputs/benefitsreport.aspx>

¹⁴ This resource is now available at: <http://datalib.edina.ac.uk/mantra/libtraining.html>



- Ekmekcioglu, C. (2007). *Research computing consultation*. University of Edinburgh. Information Services. Retrieved from <http://hdl.handle.net/10283/47>
- EPSRC. (2011). *Policy framework on research data*. Retrieved from <http://www.epsrc.ac.uk/about/standards/researchdata/Pages/policyframework.aspx>
- LIBER (2012). *Ten recommendations for libraries to get started with research data management*. Final Report of the LIBER Working Group on E-Science / Research Data Management. Retrieved from <http://www.libereurope.eu/news/ten-recommendations-for-libraries-to-get-started-with-research-data-management>
- Rice, R. & Haywood, J. (2011). Research data management initiatives at University of Edinburgh. *International Journal of Digital Curation*, 6(2), 232-244. [doi:10.2218/ijdc.v6i2.199](https://doi.org/10.2218/ijdc.v6i2.199)
- Rosenthal, D. S. H. (2010). Format obsolescence: Assessing the threat and the defenses. *Library High Tech*, 28(2), 195-210. [doi:10.1108/07378831011047613](https://doi.org/10.1108/07378831011047613)
- The Royal Society. (2012). *Science as an open enterprise*. Retrieved from <http://royalsociety.org/policy/projects/science-public-enterprise/report/>
- University of Edinburgh. (2012). Research data management roadmap. Retrieved from <http://www.ed.ac.uk/schools-departments/information-services/about/strategy-planning/rdm-roadmap>
- UKRIO. (2009). *Code of practice for research*. Retrieved from <http://www.ukrio.org/publications/code-of-practice-for-research/>