

Case Study

17

Research Data Management at the University of Edinburgh: How is it done, what does it cost?

Authors: Robin Rice (Data Librarian & Head, Research Data Support) and David Fergusson (Head of Research Services at the University of Edinburgh Information Services Group)

Email: R.Rice@ed.ac.uk

RESEARCH DATA MANAGEMENT PROGRAMME (2012-2016)

17.1. Beginnings: Data Asset Framework project

Research Data Management (RDM) began to be addressed at the University of Edinburgh in 2009, when an RDM website was published within Information Services giving advice on various RDM topics, with pointers to contacts in the University who could help. This came about because of the results of an investigation into data management practices in three university departments as part of the Digital Curation Centre's Data Asset Framework (DAF, then Data Audit Framework) in 2008.¹ Specifically, the project steering group recommended four further actions:

1. University policy in research data management;
2. Training for staff and postgraduates;
3. Web page guidance on research data management;
4. Gap analysis of existing support services.

17.2. Policy development

Another committee, led by the Director of Library & University Collections, developed the university's Research Data Management Policy, passed by the University Court in May, 2011.² This outlined the roles and responsibilities of researchers and the university itself for RDM good practice, noting that some services would need to be developed to fulfil the institution's obligations. The Data Library team began piloting training for PhD students, leading to a Jisc-funded project, Research Data Management Training (MANTRA, 2010), which created an open educational resource with 9 modules on data management and handling, still maintained and in use today.³

17.3. RDM Roadmap

The gap analysis activity, led by a new, academic-led RDM Steering Group, resulted in an RDM Roadmap, designed for high level planning that would become a living document as goals were met and new ones added. Eventually the Roadmap covered the time period January 2012 through July 2016,⁴ and covered four categories of service:

- RDM planning: support and services for planning activities typically performed before research data is collected / created;

¹ DCC: <http://www.data-audit.eu/>; last accessed 12 February 2017.

² University of Edinburgh: <http://www.ed.ac.uk/is/research-data-policy/>; last accessed 12 February 2017.

³ EDINA: <http://datalib.edina.ac.uk/mantra/>; last accessed 12 February 2017.

⁴ University of Edinburgh: <http://www.ed.ac.uk/is/rdm-roadmap/>; last accessed 12 February 2017.

- Active data infrastructure: facilities to store data actively used in current research activities and to provide access to that storage, and tools to assist in working with the data;
- Data stewardship: tools and services to aid in the description, deposit, and continuity of access to completed research data outputs;
- Data management support: awareness raising and advocacy, data management guidance and training.

17.4. Business case to fund the Roadmap

Capital and recurrent funds were secured from the university to cover the human and physical infrastructure needed to support the services. As stated in the RDM Roadmap, the business case submitted to the University IT Committee in June 2012 estimated a cost of £1M one-off, and £250K recurrent to implement the RDM Policy. In some cases, services already existed and just needed to be brought under the governance of the RDM steering group, such as Edinburgh DataShare, first set up as a demonstrator for managing research data in institutional repositories (DISC-UK DataShare project, 2007-09⁵), but becoming through completion of Roadmap goals the University's institutional research data repository.⁶ The recurrent funds enabled some new RDM-specific posts to be created, as an adjunct to existing support roles across Information Services.

SUPPORTING INFRASTRUCTURE FOR RESEARCH DATA MANAGEMENT (RDM) AT THE UNIVERSITY OF EDINBURGH

17.5. Efficiencies

The University of Edinburgh provides a consolidated underlying infrastructure for a large number of services that require data storage and data management. This has a number of advantages in both providing increased usability for the end user, all their data is accessible in 'one place' (although through numerous services), thus more closely corresponding to their mental model; and providing efficiencies of scale for the operation of the services through avoiding fragmentation and duplication of infrastructures.

17.6. Scale

The infrastructure at Edinburgh is composed of a primary base layer of a fast parallel storage file system (using GPFS), approximately 9 Petabytes in total. This allows large numbers of concurrent accesses without degrading the performance. This is then presented to a range of desktop services through a presentation layer of servers that export the correct protocols for Windows, Mac, Linux, etc. In addition, the infrastructure also serves the University compute cluster for large scale analysis tasks.

17.7. Infrastructure staffing

In order to provide this underlying converged infrastructure a small operations staff run the infrastructure, and financial support for these posts is shared across specific services (and also specific research activities and projects). There are two specific posts which support the RDM services directly on DataStore, 1 Senior Systems Engineer and 1 Junior Systems Engineer. These provide office hours support for the service and an approximately 99% availability service.

⁵ DISC-UK: <http://www.disc-uk.org/datashare.html>; last accessed 12 February 2017.

⁶ University of Edinburgh: <http://datashare.is.ed.ac.uk/>; last accessed 12 February 2017.

17.8. DataStore

The large scale shared high performance storage infrastructure at the University of Edinburgh was initiated in 2005 under the umbrella of the ECDF (Edinburgh Compute and Data Facility). The DataStore service has grown from that starting point to provide the main storage, back-up and disaster recovery infrastructure for research data, group data and personal data. Storage on DataStore is currently charged internally to the University at £175/TeraByte/year.

FROM RDM PROGRAMME TO RESEARCH DATA SERVICE

17.9. Transition and new website

The transition from RDM Programme to Research Data Service has been completed and the final Roadmap has been signed off by the steering group, with acknowledgement of some minor missed targets that will be rolled into ongoing service improvements. The new service website⁷ reflects all of the service components including DataStore and DataShare and is organised according to a vision which takes into account the full user experience of using the service in the context of doing their research, and in becoming a one-stop shop for any research data-related needs:

- User-friendly navigation and headings (instead of brand names, for example 'Active Data Storage' under the general category, 'Working with Data' instead of 'DataStore');
- Tools and support categorised according to a simplified data lifecycle corresponding to before, during, and after a research project;
- Generic and customised training and support available on demand.

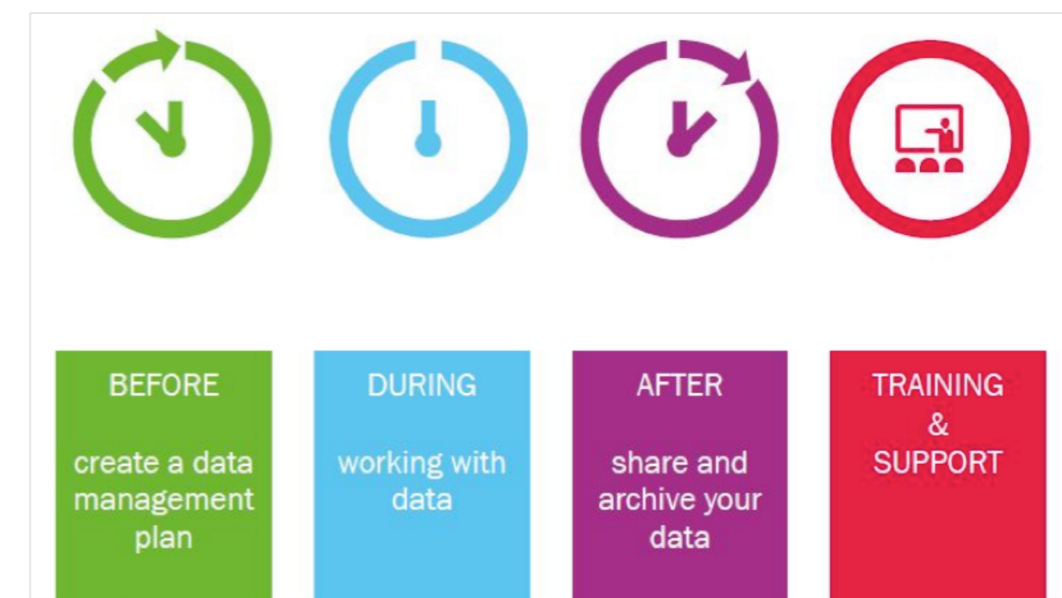


Figure 17.1: Data lifecycle and training needs

⁷ University of Edinburgh: www.ed.ac.uk/is/research-data-service; last accessed 12 February 2017.

17.10. Service team

The line is difficult to draw on the exact RDM team because of the necessary and pre-existing contributions from staff across Information Services, namely: Library & University Collections, IT Infrastructure, EDINA and Data Library, Digital Curation Centre, and User Services Division. Using service management framework language, the Research Data Service requires the following roles to be filled: Business Owner (representing the customer, currently filled by the Chair of the Steering Group), Service Owner, Service Operations Manager, and Virtual Team; the latter three are all staff members of Information Services.

17.11. Funded RDM posts

The Virtual Team is large, and includes IS staff who contribute to the service in any substantial way, and who were already providing data services of some sort before the RDM Programme began. However, the posts specifically funded by the service itself are as follows:

- 1 RDM Service Coordinator (0.7 FTE) (Library & University Collections)
- 1 Information Systems Developer (Library & University Collections)
- 1 Senior Systems Engineer (IT Infrastructure)
- 1 Junior Systems Engineer (IT Infrastructure)
- 3 Research Data Service Assistants (2.0 FTE) (EDINA and Data Library)
- 1 Software Engineer (0.5 FTE) (EDINA and Data Library)

The RDM Service Coordinator and the Research Data Service Assistants make up the front-facing staff, along with other existing posts such as data librarians.

17.12. Staff Budget

Median staff costs are 42,000 GBP for senior staff and 34,000 GBP for junior staff. The RDM budget including a small amount for operational costs (events, printing, minimal travel expenses) is 350,000 GBP for 2016-17, although due to normal staff changes and turnover this may vary some from projected expenditure. The funding model builds on the original recurrent university funds and employs cost recovery where practicable, especially via line items in research grant proposals. Hardware costs are considered capital spend.

17.13. Ongoing work

Current project activity will lead to additional service components being incorporated: Data Vault and Data Safe Haven. (Data Vault is currently being offered by appointment only but the aim is to move to a self-service workflow as soon as possible.) Data Safe Haven, an active data infrastructure for sensitive data, is due to be rolled out in August, 2017. Models to sustain their operations are being developed as part of the project activity.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.