

The International Journal of Digital Curation

Volume 8, Issue 2 | 2013

Creating a Research Data Management Service

Thomas Parsons,
ADMIRe, Jisc Managing Research Data,
University of Nottingham

Abstract

This paper provides an overview of the elements required to create a sustainable research data management (RDM) service. The paper summarises key learning and lessons learnt from the University of Nottingham's project to create an RDM service for researchers. Collective experiences and learning from three key areas are covered, including: data management requirements gathering and validation, RDM training, and the creation of an RDM website.





Introduction

Research data management is a growing area across the UK. Recent mandates by the Research Councils UK (RCUK)¹ and the growing public awareness of open access to data require a response by publicly funded research institutions. There is little doubt that compliance with and adherence to the array of rules and regulations imposed on researchers today is challenging; and is an area that requires significant support at an institutional level (Pryor, 2012).

The ADMIRE project is a JISC Managing Research Data (MRD) funded project designed to understand and address these issues, and to create a sustainable research data management infrastructure at the University of Nottingham. The overall aim was to:

‘Establish and pilot a sustainable research data management (RDM) infrastructure for the University of Nottingham. It aims to develop an infrastructure to support the research data lifecycle, acknowledging and responding to differing practices across subject disciplines’ (Parsons, 2013).

The RCUK mandates have been interpreted by funders such as the EPSRC,² so part of the work is to understand these expectations and to implement a support and technical infrastructure. This will enable:

1. Research data management throughout the research lifecycle,
2. The publication and sharing of research data.

This paper details the progress on these tasks, including the practical approaches taken, the challenges faced and the lessons learnt while creating a new RDM service at the University of Nottingham. This paper will be of value to other research institutions as they begin to address their own RDM requirements. Particular attention is paid to the elucidation of requirements for both funding and researchers, the creation of RDM training packages and the launch of RDM website to support researchers.

Requirements for an RDM Service

The creation of an RDM service is dependent upon understanding the landscape a research project inhabits. There are a variety of funding streams, both external and institutional, and a multitude of research paths and methods with which research can take. Engagement with Principal Investigators (Pis), who plan and manage research projects, has helped to understand how researchers manage data and provided a good understanding of the current levels of awareness of RDM among academics. Our approach was to instigate eight RDM pilot studies across all faculties of the university, including:

¹ RCUK common principles of data policy: <http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>

² EPSRC expectations: <http://www.epsrc.ac.uk/about/standards/researchdata/Pages/expectations.aspx>

- 
- Medicine and Health Sciences,
 - Arts,
 - Social Sciences,
 - Engineering and Science.

Interviews, surveys, focus groups and observational techniques were used to create a thorough list of candidate requirements and to understand the nature of current RDM practice. These requirements and practices were then compared to the minimum requirements stipulated by the RCUK mandates, allowing us to create a set of baseline RDM service requirements. In order to validate our thinking and ensure we were creating a sustainable service, the project utilised a survey as one of the initial steps.

RDM Survey

As with any institutional service, gaining stakeholder engagement early on is important not only to understand user requirements, but also to promote the fledgling service to academics and install RDM champions within faculties. This process was managed firstly through the pilot studies and the nomination of champions, both within the support services (Library and IT) and within the academic communities themselves. The first point of call for widespread engagement was through the use of a tailored RDM survey as part of the requirements gathering phases. The survey was designed and disseminated to academics across the university and served three purposes:

1. To gauge current levels of RDM practices,
2. To gather researcher's requirements for RDM,
3. To raise awareness for the prospective service and gauge interest levels for the proposed service.

The survey covered typical aspects of RDM and provides a benchmark to measure progress against the Research Council UK's expectations for RDM. Conducting this survey gave us a clear view of current practice and identified areas for improvement.

The survey questions were based upon the DCC's Digital Asset Framework (DAF) methodology³. Similar surveys have been carried out by Exeter, Edinburgh and Northampton using the DAF methodology, thereby allowing comparative analysis across institutions, if required. The survey instrument consisted of twenty eight questions in total. These included questions seeking demographic information, with a number of questions gathering richer data depending, upon prior answers. Questions were multiple choice (one answer), multiple choice (multiple answers possible) and free comment. Thus the analysis carried out upon completion was both qualitative and quantitative.

A small pilot group of ADMIRE project members and university researchers acted as testers for the survey design. Changes were made to the questions based upon their feedback and this ensured accuracy before sending to a wider audience. The wider

³ DCC Digital Asset Framework:

<http://www.dcc.ac.uk/resources/repository-audit-and-assessment/data-asset-framework>

survey sample consisted of career researchers (i.e. lecturers, research fellows, professors) and post-graduate researchers (those not on taught courses).

In total, 366 responses were received, with the survey running from July to September 2012. Respondents were spread across the five faculties of the university, with the larger faculties of Science, Engineering, and Medicine and Health Sciences attracting more respondents in total.

Figure 1 illustrates most common data types in order of popularity, including documents (text), spreadsheets, raw data (from software or specialist equipment), notebooks, databases and slides or specimens.

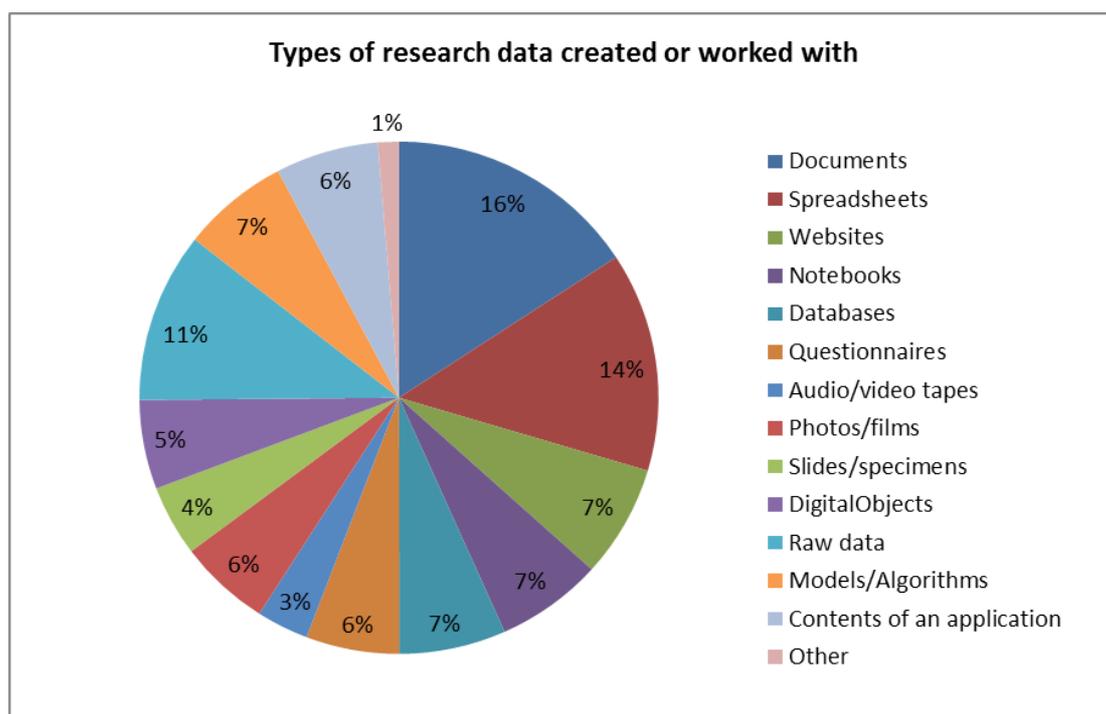


Figure 1. Types of research data created or used by survey respondents.

The data types raised an interesting divide between digital and non-digital data. The majority of researchers possessed a mixture of both types, with the digital data taking predominance in terms of volume, but the non-digital data raising complex issues of storage and identification that differed per object and type. For example, lab notebooks are widely used across the university and are typically stored in departments under the supervision of the PI. Theoretically, the EPSRC may request access to these and do stipulate that non-digital data should be converted to digital format before sharing. Conversion of data to a digital medium will therefore raise overheads and the costs of this should be factored into any proposed RDM service. This was interesting from a service perspective, as the original project scope focused upon managing and preserving digital objects rather than physical items.

In addition to the widespread types of data, the survey results indicated a widespread storage of this data. Researchers typically stored their data in at least four

places, with a significant number storing data in more than eight places and a smaller number storing data in only one to two places.

Common storage mediums in order of importance were:

- University managed computers and laptops,
- Networked university drives,
- External hard drives,
- USB pen drives,
- Web based services
- On paper.

Again this illustrates that research data management practice differs between researchers. Follow-up interviews confirmed the majority of data is left where it is generated (e.g. a lab machine hard drive) and then transferred to where a researcher feels their data is safe. In the majority of cases, this was either the university file storage or an external hard drives. However, many others are reliant upon web-based services such as Dropbox and Google Docs, thereby raising issues of data privacy, licensing and Data Protection Act compliance, primarily due to the risks associated with storing potentially sensitive data outside of the UK, as discussed in greater detail by Donnelly and Pryor (2010).

In addition to providing a baseline of current practice across researchers, the survey also provided information on what type of service should be provided. The project used the DCC lifecycle model⁴ to predict the varying training stages required by researchers. The survey results showed a wide interest in three key areas, namely:

1. Data management plans,
2. Storing data,
3. Cataloguing and metadata.

The steps take to address these areas are detailing in the following section. Further results of the survey will be released as part of the project outputs in March 2013 on the ADMIRE project blog.⁵

The results of the survey also identified that it is important to provide practical advice to researchers throughout the research data lifecycle. RDM is a relatively new area and one that can be misconceived by researchers. Work by Colombo et al. (2012) reveals that at a student level, data management training is usually restricted to information science programmes only. Our RDM survey confirmed this and found that only 7% of respondents had undertaken any form of RDM training – with the majority of these being external to the university – thereby presenting an excellent opportunity for training courses to be administered on a more structured basis. Based upon these factors, it was decided to produce training material for the two target demographics: postgraduate and career researchers.

⁴ DCC Curation Lifecycle Model: <http://www.dcc.ac.uk/resources/curation-lifecycle-model>

⁵ ADMIRE project blog: <http://admire.jiscinvolve.org/wp/>

Training

Postgraduate Training

In the summer of 2012, the university introduced Moodle to all staff and students. This provided an excellent opportunity to provision an online RDM course using this platform. An RDM training package based upon a tailored version of MANTRA⁶ was produced. Customisation included replacing Edinburgh references and adapting the text and imagery to be Nottingham-specific. Minimal customisation of the 'look and feel' and layout was required. Further work and piloting with postgraduates is planned for February to April 2013.

Researcher Training

In many respects, the provision of career researcher training is more pressing, as this is the group who create, write and submit the research proposals that generate income for the university and, consequently, fund the majority of postgraduate researchers. As PIs, they are responsible for meeting the mandates set out by the funding councils, and more importantly, for writing a data management plan to support their funding applications. The survey results of Figure 2 illustrate the type of training requested by all researchers:

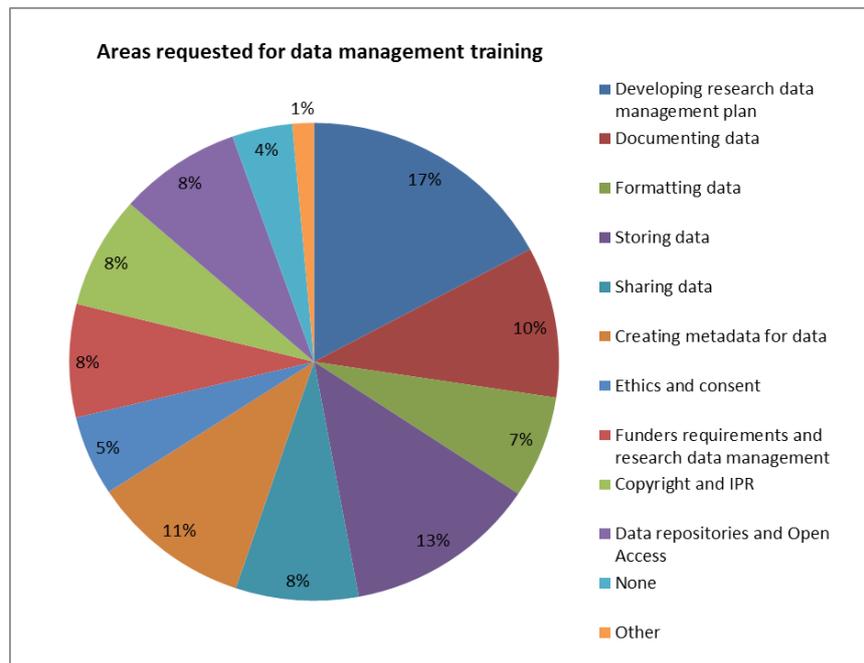


Figure 2. Data management training areas requested by survey respondents.

As is to be expected, further analysis of respondents based upon seniority found the areas of practical concern featured the strongest responses. Career researchers must know how to create a data management plan (DMP). Follow-up focus groups suggest

⁶ Research Data MANTRA (online course), produced by EDINA and the Data Library, University of Edinburgh: <http://datalib.edina.ac.uk/mantra>

the majority are unclear as to the level of detail and guidance required, a finding echoing that of Ward et al. (2011). While all researchers recognised that learning about data storage would be advantageous to their work, less are interested in the practices of sharing data. In many respects this indicates the infancy of RDM within Nottingham and the research sector as a whole. If the majority of researchers had already completed a DMP that stated they are going to share their data at project end, then the percentage requiring advice on where and how to share data would almost certainly be higher.

Likewise, we found there is a healthy interest in metadata training, perhaps related to the survey finding that only 7% of respondents currently assign metadata to their data. Of these, the majority use proprietary or subject-specific standards, so the ADMIRE project is undertaking steps to produce a dedicated training package for ‘tagging’ or assigning metadata to data. As Groenewald and Breytenbach (2011) note, preserved digital objects will only have meaning to others when they are accompanied by descriptive, structural and technical (administrative) metadata – elements and activities which must be explained and simplified when designing training that will be given to researchers of differing RDM knowledge and expertise.

Delivery of RDM training will be via:

- Online courses (Moodle-based),
- Pilot DMP workshops per faculty (e.g. Medicine and Health Sciences will have MRC⁷ and BBSRC⁸-focused sessions),
- Short courses run by the central university training teams.

The first and final training elements acknowledge that RDM training sessions must be available beyond the life of the ADMIRE project in order for the service to be carried forward.

In addition to specific training courses, it was clear from the outset that the service must deliver an RDM-focused website to augment regular training sessions, given the number of researchers at the university. The number of researchers at the university is simply too high to reach all researchers via face-to-face training sessions. In light of this, the project team believe that sufficient information can be provided via RDM specific webpages. Focus groups with researchers in the university suggest that the interpretation of the various policies and guidance around research data is bewildering for many researchers, so more subject-specific RDM training to run alongside the website is currently in development.

Structuring an RDM Website

As mentioned previously, the website will act as both a centralised point for information to researchers and as a public facing showcase of research at the University of Nottingham. The approach taken to designing the website was based

⁷ Medical Research Council (MRC) data policy:

<http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/datasharing/DMPs/index.htm>

⁸ Biotechnology and Biological Sciences Research Council (BBSRC) data policy:

<http://www.bbsrc.ac.uk/publications/policy/data-sharing-policy.aspx>

upon the requirements of the researchers themselves and in response to the funding bodies' emphasis on certain areas.

One of the first steps in creating a service-driven website is to identify the end users of the site and those who will support the area. In this case, the site and content is aimed at postgraduates and researchers. While the majority of information is generic and applicable to both groups, specialist advice relating to funding, DMPs and funding requirements targets only career researchers.

It should be noted that simply choosing the audience for the site required significant liaison with departments and researchers across the university. Funding information and funding requirements information was created by Research and Graduate Services (RGS) and enhanced by members of the project team, while technical information relating to data storage and provision was provided by IT Services. Finally, the university libraries provided information on copyright, metadata and licensing issues. Areas that fell between current services – such as file formats, digital archiving and interoperability – fell to the project team to write and approve. In all there are over fifty pages within the site, with the core areas covering:

- What is research data?
- The research data lifecycle
- Data management planning
- Creating data
- Organising and storing data
- Data sharing and data archiving
- Research data showcase
- Training, advice and support
- Contact us

These areas are typical of those found on other RDM sites,⁹ but include sections that define what 'research data' constitutes and a section that allows researchers to openly highlight and share their data. The 'research data showcase' is intended to not only publicise good research practice at the university, but will also shed light on what happens to publicly funded research. Given that the funding mandates will increase the amount of data available, highlighting good practice and reuse of data is required – particularly if research data will be shared with no clear view of who will reuse or even be interested in it. This point is also raised by Borgman (2011) who calls for more research into who will reuse research data, with the emphasis being to understand fields that share data freely at present and then seek to apply this learning to other areas.

The University of Nottingham RMD web site¹⁰ was launched in March 2013 and has certainly helped to promote research data management as the site has become

⁹ For example, see the University of Cambridge RDM website: <http://www.lib.cam.ac.uk/dataman/>

¹⁰ The University Of Nottingham RDM website: <http://www.nottingham.ac.uk/researchdata/>



more visible within the research community. The site launched with two research data showcase articles and there is interest from other researchers who would like to contribute to this area. It is telling that the most vocal advocates are from the biosciences and medical fields – areas that are under significant pressure to present public benefit of their work.

The act and processes of engaging researchers throughout the web site content and development phases allowed the project to highlight the gaps in the current service and in many ways, was viewed as analogous to the creation of the new service. Libraries, IT Services and RGS are heavily involved in production of the web site content and as such, have tentative claims of ownership of their respective areas in the absence of a central RDM team. In light of the multi-disciplinary nature of the service, the site location and URL were purposely chosen so as to serve the researchers and not indicate ownership at this stage in the work. Rice and Haywood (2011) note cross departmental working is necessary when creating such a service, with Edinburgh utilising both library and IT staff to form an RDM support group.

This final point leads to an important number of lessons learnt for other institutions on the road to creating an RDM service.

Lessons Learnt

Learning from experience is invaluable to projects of this nature and the publication of experiences via JISC MRD workshops, blogs¹¹ and reports has been invaluable. Our experiences as RDM service practitioners has been captured as we have progressed through the project. We have focused on important topics such as gaining institutional engagement, positioning an RDM service within a university, roles and responsibilities, and the lessons learnt from our series of pilot studies. Of these the following areas were deemed critical.

Senior Management Buy-In

There is little doubt RDM is a high profile service that is driven by those who fund research. Yet, although the mandates are in place, there is little explicit evidence of sanctions against those who do not comply or evidence of reduced funding as a result of non-compliance. The majority of mandates make it clear that penalties may occur should projects not actively manage research data throughout the project lifecycle or share data post-project. This potentially affects all aspects of the university and therefore requires senior steer and support at Pro-Vice-Chancellor level and above. In the case of the ADMIRE project, a steering group at this senior level will take ownership of the service at the end of the project. The value of this group has already been demonstrated by having the authority to take the operational and resourcing decisions necessary for the service, and the gravitas to drive a draft RDM policy through to implementation.

¹¹ JISC MRD Orbital blog: <http://orbital.blogs.lincoln.ac.uk/>



Technical Infrastructure

This paper has purposefully not discussed the technical elements that need to support an RDM service, but rather focused on the ‘softer’ sides of RDM. However, the funding mandates suggest that data be kept safe, secure and accessible (whether publicly or not) for up to ten years.

Scaling-up and purchasing new software and hardware to meet these demands is dependent upon strong cost benefits and clear drivers, given the level of infrastructure investment required. We may surmise that that the drivers for this investment are in place, but the cost benefits are less clear, particularly given that articles such as Lord and Macdonald (2003) suggest multiple levels of curation, archiving and preservation are required – all of which fall outside the scope of current university IT services. Therefore, when undertaking such a project, it is often simpler to focus on the ‘low hanging fruit’ or human aspects of the service. Yet, as in our case, simply increasing awareness of RDM to researchers through the use of a survey generated a large number of requests for the technical infrastructure, before this has been thoroughly scoped or even piloted. However, the demand from researchers has accelerated the development of the technical infrastructure and demonstrated the need for investment in both the technical and support infrastructure.

A Long Term versus Short Term Solution

As this project has a limited life, there have been conflicts between delivering for the project and delivering what will be a sustainable service long-term. In many ways, the project has chosen the long-term path by creating an RDM website, training materials and promotional materials that match the branding of the university. However, the immediate demand for these services, particularly with regards to creating DMPs, has created uncertainty with regards to responsibilities of project and permanent university staff, particularly with regards to who will support aspects of the service and who should engage with researchers. Again, this highlights that an initial scoping exercise can quickly manifest into a service that is resourced only by temporary project team members. We quickly learnt that a pilot service should be backed up by permanent staff with sufficient time to provide their services.

Conclusion

In conclusion, the field of RDM and this project is at a turning point. The initial scoping and requirements phases of this project have resulted in a strong demand for the service from both researchers and university support staff. Likewise, RDM is a growing field of research in itself and, as discussed, the technical or service architectures that will meet funding requirements are well documented; yet the practical interpretation of what the policies mean and their use by researchers on a daily basis is not.

Therefore, the remaining work for this project is to utilise the mandates and researcher’s enthusiasm to launch the service in 2013 as a pilot, with the view to moving to a resourced and sustainable service by the end of 2013, thereby mirroring developments in other research-led universities and hopefully meeting funding mandates with as little disruption to the researcher’s daily work as possible.



References

- Borgman, C. (2011). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*.
[doi:10.2139/ssrn.1869155](https://doi.org/10.2139/ssrn.1869155)
- Colombo, P., Piorun, M., Kafel, D., Leger-Hornby, T., Najafi, S. & Martin, E. (2012). Teaching research data management: An undergraduate/graduate curriculum. *Journal of eScience Librarianship* 1(1). [doi:10.7191/jeslib.2012.1003](https://doi.org/10.7191/jeslib.2012.1003)
- Donnelly, M. & Pryor, G. (2010). The fourth DCC-RIN research data management forum. *Ariadne* 63. Retrieved from <http://www.ariadne.ac.uk/issue63/rdmf4-rpt/>
- Groenewald, R. & Breytenbach, A. (2011). The use of metadata and preservation methods for continuous access to digital data. *The Electronic Library*. 29(2), 236-248. [doi:10.1108/02640471111125195](https://doi.org/10.1108/02640471111125195)
- Lord, P. & Macdonald, A. (2003). *e-Science curation report – Data curation for e-Science in the UK: An audit to establish requirements for future curation and provision*. Prepared for the JISC Committee for the Support of Research (JCSR). Twickenham, UK: The Digital Archiving Consultancy Limited. Retrieved from: http://www.jisc.ac.uk/uploaded_documents/e-ScienceReportFinal.pdf
- Parsons, T. (2013). *JISC ADMIRE project plan*. Retrieved from <http://admire.jiscinvolve.org/wp/files/2013/05/ADMIRE-Project-Plan-Public-May-2013.pdf>
- Pryor, G. (Ed). (2012). *Managing research data*. London, UK: Facet Publishing.
- 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)
- Ward, C., Freiman, L., Jones, S., Molloy, L. & Snow, K. (2011). Making sense: Talking data management with researchers. *International Journal of Digital Curation* 6(2), 265-273. [doi:10.2218/ijdc.v6i2.202](https://doi.org/10.2218/ijdc.v6i2.202)