



Leading from the library: data management initiatives at the University of Northampton

Miggie Pickton,

Research Support Librarian,
University of Northampton,
Northampton, England, UK
E-mail: Miggie.Pickton@northampton.ac.uk

Sarah Jones,

Research Officer,
Digital Curation Centre at HATII,
University of Glasgow, Scotland, UK
E-mail: Sarah.Jones@glasgow.ac.uk

Marieke Guy,

Research Officer,
Digital Curation Centre at UKOLN,
University of Bath, England, UK
E-mail: m.guy@ukoln.ac.uk

Session:

116 — *The role of libraries in data curation, access and preservation: an international perspective* — Science and Technology Libraries

Abstract :

This paper reflects on data management initiatives at the University of Northampton and the central role of the library in pushing forward this work. Over the last two years the University has implemented several tools provided by the UK Digital Curation Centre (DCC) and is now collaborating directly with the DCC to implement its new research data policy and train librarians to provide research data management (RDM) support. The DCC is engaging with several universities in this way so sets the example of the University of Northampton in a wider UK context.

Introduction

The digital age affords new opportunities and researchers are being encouraged to open up their data to capitalise on them. Developments in technology and techniques allow new research questions to be answered. Vast quantities of data can now be created, processed and shared, allowing 'grand challenges' to be addressed on a global scale. Such opportunities are premised on having well-managed data that are appropriately documented to be understandable to others, underlining the importance of good research data management (RDM) practice.

Developments in legislation and research governance have placed research data management firmly on institutional agendas. Several prominent Freedom of Information (Fol) requests for research data have had a detrimental effect on institutional reputations and have brought the risks of poor data management into sharp focus. In the same time period, changes in overarching research governance have clarified the role universities should play in supporting the effective management and sharing of research data. The driving principle of data as a public good has led to requirements for greater transparency and demonstration of research integrity.

With the changing landscape in mind, many Higher Education Institutions (HEIs) are developing policies and implementation plans for research data management. The Engineering and Physical Sciences Research Council (EPSRC) policy framework for research data, which was released in May 2011, has been a key impetus for many universities in this area. The policy sets out clear timescales for implementation: the EPSRC expects research organisations that it funds to have developed a roadmap for RDM by 1st May 2012, and to be fully compliant with these expectations by 1st May 2015 (EPSRC, n.d).

About the DCC

The Digital Curation Centre (DCC) was established as a national centre to solve challenges in digital curation that could not be tackled by any single institution or discipline. It launched in 2004 and is now in its third phase with team members based at the Universities of Edinburgh, Bath and Glasgow. Work activities have evolved during its three phases and there has been a recent shift away from the development of curation tools to building capacity, capability and skills for data management across the UK's higher education research community. This new emphasis has exposed a critical dependency upon the contribution of a network of practitioners beyond the core DCC, who will be crucial to the exponential growth of effective RDM practice.

Over the 2011/2013 period the DCC has been funded by the Higher Education Funding Council for England (HEFCE) under the Universities Modernisation Fund (UMF) to assist universities in developing research data management strategies and services. It is currently undertaking in-depth engagements with eighteen HEIs with a view to increasing data management capabilities and developing a redeployable model for future engagements with other UK universities. One of the institutions in the DCC engagement programme (DCC, n.d.) is the University of Northampton, where RDM activities are led by the Research Support Librarian.

About the University of Northampton

The University of Northampton is a relatively new university having only been awarded research degree awarding powers in 2005. In achieving university status the institution's ambition was to become a leading regional, national and international centre for research and knowledge transfer. Thus the period that followed saw an increasing focus on research and, from the library's perspective, a series of initiatives to support researchers.

In 2007, an organisational restructuring of library and IT services resulted in the creation of a dedicated post for research support. The library based post holder was tasked with developing services for researchers across the university. One of the first developments was an institutional repository, NECTAR (the Northampton Collection of Theses and Research).

The library was fortunate in receiving funding from the JISC (<http://www.jisc.ac.uk/>) to support this work. The JISC also supported a further project, the KeepIt project (Hitchcock, 2010a) which explored issues around the preservation of repository content. It was during the KeepIt training course that library staff were first introduced to some of the DCC data management tools and a role was identified for librarians in using and supporting these (Hitchcock, 2010b).

The role of the librarian in supporting research data management

Northampton is not alone in having librarians lead the way in promoting and embedding good research data management practices. Many of the DCC engagements are being carried out in conjunction with research support and library service teams. Half of the engagements have been initiated by the library, while many others are being steered by research offices with the library undertaking much of the work. Only 2 engagements out of 18 are led by IT services.

Indeed, data management initiatives at several UK universities have emanated from the library. At the University of Edinburgh, for example, the data library at EDINA has been pivotal in initiating activity and sustaining momentum (Rice and Haywood, 2011). Meanwhile the Bodleian Library at the University of Oxford is playing a key role in the long-term management, publication and reuse of research data through its DataBank repository (<https://databank.ora.ox.ac.uk/>), and JISC Managing Research Data (MRD) projects at the University of Southampton have put forward the role of subject librarians in supporting research data management (JISC, 2012a; JISC 2012b). These moves parallel activity in other countries where librarians have naturally taken on the responsibility to support RDM, for example in North America where there is a strong iSchool tradition (Hswe and Holt, 2010).

In the UK, university libraries typically host the institutional repository for access to research publications. As such, they are often the stakeholder called on when questions are raised about the management of associated data. The information science skills held by librarians and archivists are key constituents for research data management. The 2012 RLUK report on Re-skilling for Research (Auckland, 2012) looks at the role and skills of subject and liaison librarians required to effectively support the evolving information needs of researchers. It states that:

“it is clear that as the nature of research within our institutions changes, so must the role of the library in supporting research”

(David Prosser in Auckland, 2012, preface)

Data is gaining in significance as a research output and is increasingly expected to be published and shared (Borgman, 2011). Librarians have directed the open sharing of publications so are well placed to advise on how best to support data requirements. The inclusion of library expertise in institutional working parties and

steering groups bent on designing the institutional strategies and infrastructure necessary to meet new operational and regulatory requirements is paramount.

Developing a research data policy for the University of Northampton

In 2010, when library staff were first introduced to the DCC tools, little was known centrally about university researchers' data storage requirements, or indeed the research workflow that incorporated the creation and management of data. There was no university wide data storage policy at Northampton, nor a recognised procedure for managing research data.

Being sensitive to the changes in the research landscape described above, senior research managers welcomed the library's proposal to conduct a Data Asset Framework (DAF) project. DAF is a methodology which enables universities to audit research data collections as well as awareness, policies and practice for data curation and preservation. It was therefore ideal for gaining an understanding of existing research data management practices at Northampton and providing an evidence base to support future data management initiatives.

The DAF methodology comprises four steps:

1. "Stage 1 is for planning, defining the purpose and scope of the survey and conducting preliminary research.
2. Stage 2 is about identifying what data assets exist and classifying them to determine where to focus efforts for more in-depth analysis.
3. Stage 3 is where the information life cycle is considered to understand researchers' workflows and identify weaknesses in data creation and curation practices.
4. Stage 4 pulls together the information collected and provides recommendations for improving data management."

(DCC, 2009, p.5)

At Northampton, the project team comprised the library's Research Support Specialist and two project researchers (both graduate interns). The team was advised by a Project Board composed of staff with expertise in repositories, records management and collection development. The project ran for eight weeks from May to June 2010.

The aims of the project were three-fold:

1. To investigate the types of data held by researchers throughout the university, researchers' existing data management practices and the risks associated with these practices.
2. To provide evidence to inform a possible new data management policy and services to satisfy the requirements of researchers and funders.
3. To raise awareness amongst researchers about good data management practice, including the provision of short and long term data storage and access.

Following the four steps of the DAF methodology, the project team first researched other implementations of the DAF methodology and defined the overall scope and outline of the project. With this knowledge, it was decided that data would be collected in three ways: interviews with research leaders in each of the six Schools;

an online survey of researchers; and in-depth follow-up interviews with researchers who volunteered having completed the survey.

The interviews with research leaders gave the project team a broad understanding of the different disciplinary data management practices and the chance to achieve 'buy-in' from influential researchers in each School. The online survey, which was open to all researchers in the institution and attracted 80 responses, covered topics such as the types, sizes and formats of research data held; its ownership; means of storage; security arrangements; sharing and access over the short and long term; and the requirements of funders. The final interviews with 16 researchers enabled the project team to follow up key findings from the survey and gather additional technical information on specific data objects.

The findings of the project are described in the project report (Alexogiannopoulos *et al.*, 2010). A number of themes emerged.

Three generic types of researcher were identified, based on the different needs and behaviours they demonstrated with respect to research data: the research student, the independent researcher and the group researcher/collaborator. Some common behaviours were identified, for example, researchers overwhelmingly used Microsoft software for creating documents and spreadsheets and so habitually created .doc and .xls file types; similarly, .jpeg was the preferred format for image files. In contrast, there was much greater variation in the file types used for databases, audio and video files.

These findings have significant implications for data curation. Data storage needs and behaviours appeared to vary throughout the research lifecycle, with different storage devices being prominent at the data collection, analysis and project completion stages. For those that needed to share data, access to a shared server was effective, but where this was not available, email was most frequently used. Very few Northampton researchers had at that time applied for funding from a body that mandated open access to research data and only just over half were interested in a university repository for data (either for open or closed access).

The DAF project report made a number of recommendations, among them the creation of a university research data policy. Upon receipt of the report the University Research Committee convened a Research Data Working Group. The Working Group comprised representatives from the research community, records management and the library and was chaired by the library's Research Support Specialist.

The Working Group discussed the potential scope of the policy, its fit with the research lifecycle and its relationship with other university policies (for example, research ethics and academic misconduct policies). The Research Committee had also made it clear that the policy should be accompanied by a viable supporting procedure so this was also considered.

Having reviewed both funders' research data policies, framed by the Research Councils UK (RCUK) Common Principles on Data Policy (RCUK, n.d.) and those of other institutions (such as that of the University of Melbourne (2010)), the Working

Group proposed a policy in which researchers were expected to follow key RCUK recommendations for record-keeping, data security and open access (RCUK, 2009, p.5). To facilitate this, researchers would be expected to complete a Data Management Plan (DMP) at the start of each project, the university would be asked to provide a dedicated storage facility for research data, and library and records management staff would provide appropriate training and support.

There are a number of tools and checklists available for data management planning, for example DMPTool (<https://dmp.cdlib.org/>), DMPonline (<https://dmponline.dcc.ac.uk/>) and the Australian National Data Service data management planning checklist (<http://ands.org.au/guides/data-management-planning-awareness.html>). The Working Group chose DMPonline as its preferred tool. DMPonline was originally developed by the DCC to help researchers respond to the data management requirements of the major RCUK funders when submitting grant proposals. However, at Northampton it was felt that there was a greater need for a more generic tool; one which supported good practice in data management irrespective of who was funding it, or even of whether the research was externally funded at all. Once again, librarians led the way, liaising with DCC staff to develop a new template in DMPonline.

The Working Group's first proposal for a research data policy was presented to University Research Committee in January 2011. Members of the committee expressed some concerns. It was felt that the policy was too prescriptive; it was not relevant or applicable to all disciplines and in some cases it necessitated duplication of effort. A revised policy, which retained the same expectations but encouraged rather than mandated the creation of DMPs and proposed a less formal monitoring process, was approved by University Research Committee in June 2011 (University of Northampton, 2012).

Implementing Northampton's research data policy

The existence of a policy does not in itself alter behaviour. Implementation requires appropriate infrastructure, enthusiastic advocacy and the leadership of early adopters. These challenges are now being addressed with DCC support.

Much of the infrastructure is in place, so the primary step for Northampton is to embed good research data management across the institution. DCC and library staff will work with a pilot group of researchers to test the practical implementation of the policy. The researchers' experiences will form the basis of training and advocacy materials to roll the policy out further. A bid has been submitted to the training strand of the JISC MRD programme (JISC, 2012c) to support this work. If successful, a number of additional deliverables will be produced, including the provision of a set of exemplar case studies.

It is important that the pilot group represents a wide range of research scenarios, covering different disciplines and cross-disciplinary collaboration. Ideally it will include members of all six university Schools. It should involve funded and non-funded work by both new and established researchers, working as individuals and in groups. The participating researchers should create a diverse range of data types

and formats, to allow deposit processes for both open and closed datasets to be tested.

The use of a pilot group will serve two purposes. Participating researchers will represent the views and research data management needs of their own disciplines (and thus help to identify gaps in provision to develop relevant training materials) and the resultant outputs will exemplify the potential benefits of good RDM practice and demonstrate how these can be realised by other research projects. The Data Management Plans to be created by early adopters, for example, will act as model for others to repurpose.

The practical experience of supporting RDM gained through the pilot group will provide content for a second set of training materials aimed at university librarians, research managers and administrators, IT staff and records managers who support research data management. These materials will draw on existing resources such as the DCC's DC101 course materials (DCC, 2008) and Data Intelligence 4 Librarians by the 3TU consortium in the Netherlands (3TU.DataCentrum, n.d.). The mode, format and duration of training will reflect the needs of the various audiences but is likely to include hands-on workshops, online tutorials and shorter talks to raise awareness.

Feedback from the pilot group will also prove invaluable to help refine existing services. In terms of infrastructure development, the main task remaining at Northampton is developing a central data storage facility. The nature of this is yet to be defined, but will be informed by the work of projects such as Dataflow at the University of Oxford (<http://www.dataflow.ox.ac.uk/>) and MaDAM at the University of Manchester (The John Rylands University Library, n.d.) . This is a rapidly moving field and options for local, networked and cloud-based storage will need to be considered.

Setting the University of Northampton in the UK context

Despite being a post-1992 university, the University of Northampton is ahead of the curve in terms of research data management. While many institutions are only now beginning to investigate current practice to define their RDM strategy, Northampton has been undertaking this work since 2010. Indeed, a specific library post to support research has been in place since 2007.

This pattern is reinforced when we compare Northampton with two other post-1992 universities with which the DCC is engaging. In both of these cases, the focus of activity to date has been to undertake DAF assessments and assist in the definition of RDM policy. The only other institution at which the DCC is assisting with policy implementation is the University of Edinburgh, an older, research-intensive Russell Group university, which has been leading developments in the RDM field for some years.

Being a smaller, less research-intensive university arguably makes Northampton well-placed to be innovative. Decisions can be made quickly and researchers seem more willing to trial new initiatives - resistance to central services and fears of losing

autonomy have not proved to be stumbling blocks, as in the case of some larger universities.

Weighed against this, the smaller amount of internal resource available forces those responsible for supporting research to be creative in seeking external support – be it project funding, internships or institutional engagements. The challenge then is to ensure that knowledge transfer is permanent, by involving and up-skilling as many university staff as possible.

Conclusion

The University of Northampton is fulfilling its ambition to become a leading regional, national and international centre for research and knowledge transfer. The activities documented in this paper demonstrate its sustained commitment to supporting research over several years. Research data management needs have been identified and addressed, and support is being embedded into researchers' workflows to improve practice. The pivotal role of the library in these initiatives is indicative of broader trends and underlines the wider implications of research data management for the sector.

In contrast to fears that digital transformations are threatening the future of libraries (Lombardi, 2000), research data management highlights the applicability of the librarian's skillset. RDM offers a new role for libraries: namely to support researchers to create content, to share their data to gain academic impact, and to understand and reuse others' datasets. Librarians have a long history of providing such support for print media, so have the transferable skills necessary to support research data management. Moreover, librarians are able to engage with researchers at a fundamental level, particularly academic liaison and subject librarians given their specialisms.

The role of the library is being transformed at the University of Northampton and many other UK universities with which the DCC is engaging. Librarians are providing RDM support via guidance, training and outreach programmes, and are co-developing technical infrastructure to manage and share all types of research output. Effective management and sharing of research data, as well as publications, demonstrates quality research and underpins institutional prestige. By embracing the need to provide RDM support, librarians will remain at the heart of institutional agendas.

References

- 3TU.DataCentrum (n.d.) *Data intelligence* [online]. Available from <http://dataintelligence.3tu.nl/en/home/> [accessed 26.04.12].
- Alexogiannopoulos, E., McKenney, S. and Pickton, M. (2010) *Research Data Management Project: a DAF investigation of research data management practices at The University of Northampton* [online]. Northampton: University of Northampton. Available from: <http://nectar.northampton.ac.uk/2736/> [accessed 26.04.12].
- Auckland, M. (2012) *Re-skilling for research* [online]. Available from: <http://www.rluk.ac.uk/files/RLUK%20Re-skilling.pdf> [accessed 26.04.12].
- Borgman, C.L. (2011). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*. [online] 1-40 Available from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1869155 [accessed 26.04.12].
- DCC (2008) *DCC Digital curation 101 workshop* [online]. Available from: <http://www.dcc.ac.uk/events/workshops/dcc-digital-curation-101-workshop> [accessed 26.04.12].
- DCC (2009) *Data Asset Framework: Implementation guide* [online]. Available from http://www.data-audit.eu/docs/DAF_Implementation_Guide.pdf [accessed 26.04.12].
- DCC (n.d.) *Institutional engagements* [online]. Available from <http://www.dcc.ac.uk/community/institutional-engagements> [accessed 26.04.12]
- EPSRC (n.d.) *EPSRC policy framework on research data* [online]. Available from: <http://www.epsrc.ac.uk/about/standards/researchdata/Pages/default.aspx> [accessed 26.04.12].
- Hitchcock, S. (2010a) *KeepIt: Kultur, eCrystals, EdShare (and NECTAR) – preserve it!* [online]. Available from: <http://preservation.eprints.org/keepit/> [accessed 26.04.12].
- Hitchcock, S. (2010b) KeepIt course: revision, conclusion and evaluation. 30 September 2010. *Diary of a repository preservation project* [online]. Available from: <http://blogs.ecs.soton.ac.uk/keepit/2010/09/30/keepit-course-revision-conclusion-and-evaluation/> [accessed 26.04.12].
- Hswe, P. and Holt, A. (2010) *Guide for research libraries: the NSF data sharing policy* [online]. Available from: <http://www.arl.org/rtl/eresearch/escien/nsf/index.shtml> [accessed 26.04.12].
- JISC (2012a) *DataPool* [online]. Available from: http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata/infrastructure/datapool.aspx [accessed 26.04.12].
- JISC (2012b) *IDMB: Institutional data management blueprint* [online]. Available from: <http://www.jisc.ac.uk/whatwedo/programmes/mrd/rdmi/idmb.aspx> [accessed 26.04.12].
- JISC (2012c) *Managing research data programme 2011-13* [online]. Available from: http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata.aspx [accessed 26.04.12].

The John Rylands University Library (n.d.) *MaDAM* [online]. Available from: <http://www.library.manchester.ac.uk/aboutus/projects/madam/> [accessed 26.04.12]

Lombardi, J. V., (2000) Academic Libraries in a Digital Age. *D-Lib Magazine* [online]. **6** (10). Available from: <http://www.dlib.org/dlib/october00/lombardi/10lombardi.html> [accessed 26.04.12]

RCUK (n.d.) *RCUK Common principles on data policy* [online]. Available from: <http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx> [accessed 26.04.12].

RCUK (2009) *RCUK Policy and code of conduct on the governance of good research conduct* [online]. Available from: <http://www.rcuk.ac.uk/Publications/researchers/Pages/grc.aspx> [accessed 26.04.12].

Rice, R. and Haywood, J. (2011) Research data management initiatives at University of Edinburgh. *The International Journal of Digital Curation* [online]. **2** (6), 232-244. Available from: <http://www.ijdc.net/index.php/ijdc/article/view/194/259> [accessed 26.04.12].

University of Melbourne (2010) *Policy on the management of research data and records* [online]. Available from <http://www.unimelb.edu.au/records/research.html> [accessed 26.04.12].

University of Northampton (2012) *Research data policy* [online]. Available from <https://www.northampton.ac.uk/info/20283/academic-research/1606/research-data-policy> [accessed 26.04.12].