RESEARCH DATA MANAGEMENT AT THE UNIVERSITY OF WARWICK: RECENT STEPS TOWARDS A JOINED-UP APPROACH AT A UK UNIVERSITY

from Jenny Delasalle

Abstract

This paper charts the steps taken and possible ways forward for the University of Warwick in its approach to research data management, providing a typical example of a UK research university's approach in two strands: requirements and support. The UK government approach and funding landscape in relation to research data management provided drivers for the University of Warwick to set requirements and provide support, and examples of good practice at other institutions, support from a central national body (the UK Digital Curation Centre) and learning from other universities' experiences all proved valuable to the University of Warwick. Through interviews with researchers at Warwick, various issues and challenges are revealed: perhaps the biggest immediate challenges for Warwick going forward are overcoming scepticism amongst researchers, overcoming costs, and understanding the implications of involving third party companies in research data management. Building technical infrastructure could sit alongside and beyond those immediate steps and beyond the challenges that face one University are those that affect academia as a whole. Researchers and university administrators need to work together to address the broader challenges, such as the accessibility of data for future use and the reward for researchers who practice data management in exemplary ways, and indeed it may be that a wider, national or international but disciplinary technical infrastructure affects what an individual university needs to achieve. As we take these steps, universities and institutions are all learning from each other.

Introduction

The approach of the University of Warwick towards research data management took two main strands in the period covered in this article: setting requirements and providing support. This paper describes progress made in those areas, up until May 2013, and introduces important aspects of UK government and research funder imperatives. These imperatives added to both the challenges but also the solutions, for the University of Warwick. In this article, we see some of the steps already taken and consider the challenges and possible solutions going forward, based on the issues that arose from interviews with researchers and advice received from other UK data management initiatives.

Background

Research is a strong feature of the University of Warwick: it is one of the Russell Group Universities in the UK, ranked 7th for its research in the UK's last Research Assessment Exercise in 2008. Warwick's researchers produce around 4000 publications per year, as evidenced in the "Publications Service" section of its institutional repository (University of Warwick, 2013). Warwick is naturally interested in research data

management, as good research practice but there were other drivers for developing a more formal approach.

UK government interest in open access to publicly funded research data was signalled by a much anticipated White Paper on Open Data, published in June 2012 (HM Government Cabinet Office, 2012) and by the 'Common Principles on Data Policy' published by the Research Councils UK (RCUK, 2012). In particular, one of the most influential RCUK funding bodies, the Engineering and Physical Sciences Research Council (EPSRC) required that institutions be compliant with its Policy Framework on Research Data by May 2015 (EPSRC, 2013). EPSRC also required Warwick and all universities in receipt of its funding to develop a roadmap detailing how this compliance would be reached, by May 2012.

Beyond UK government funding drivers, the Royal Society's 'Science as an open enterprise' initiative (The Royal Society, 2013) gave Warwick's open access and data management work further impetus, as it called for open data to be the default rather than the exception, for research assessment to include data as well as publications, and for universities to invest in data training and infrastructure.

These funders' expectations provided one solution, in that they prompted attention from researchers and university managers, on the matter of research data management. The funders also backed national initiatives and brought deadlines and their own requirements that had to be met and incorporated into the University's own policy. The co-operation across the UK's national funding bodies was helpful: if funders from different disciplines had wildly different expectations then this would perhaps have been more of a challenge to incorporate at the University level.

In the UK there were also various data management focussed initiatives that provided good examples and support that the University of Warwick could draw upon. In particular, the Digital Curation Centre (DCC)[1], a national UK service centre based at the University of Edinburgh and launched in 2004, was able to support Warwick through its guidance and tools for the long-term preservation and use of research data. Lessons were already being shared by recipients of funding through JISC's Managing Research Data Programme [2] and many of Warwick's researchers were already interacting with the UK Data Archive, who "acquire, curate and provide access to the UK's largest collection of social and economic data" for instance (UK Data Archive, 2013), and more about useful examples and existing practices is revealed in the latter half of this paper.

Bring in a research data management policy: set your requirements

The first step in setting requirements was to bring in a policy on data management that made these clear: in November 2011 the University of Warwick's Steering Committee approved the University of Warwick Research Data Management Policy that was published in December 2011, as part of a Research Code of Practice (University of Warwick, 2011).

[1] See: http://www.dcc.ac.uk/

[2] See: http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata.aspx

One of the important things to go into such a policy is a definition of research data: at Warwick this is fairly broad, including inputs as well as outputs. It was also important that the policy be compatible with the requirements of important funders, and that it fit the particular needs of the University, setting the direction for the future. The librarian was very much involved in the development of the policy, along with senior management at the University of Warwick, and he was able to look at others' policies and to discuss draft policies with other institutions.

Approval of the policy was a very important first step: the steering committee at Warwick is attended by those at the very highest level at the University, including the Vice-Chancellor and Deputy Vice-Chancellor, the Pro-Vice-Chancellors (PVCs) including those for Research, the Chairs of the Boards of the Faculties, and the Chair of the Board of Graduate Studies. Such key people had set the direction for data management at the University of Warwick, by the end of 2011.

Warwick's Research Code of Practice containing the Data Management Policy is referred to by all researchers at the University: it is a particularly appropriate place for the policy to belong. The Research Code of Practice is referred to by researchers and used by Research Ethics Committees: these are a scrutiny process that many research projects, especially the larger ones, go through in order to "safeguard researchers conducting the study" and also to protect "the rights, safety, dignity and well-being of research participants" (University of Warwick, 2012).

Warwick's Research Ethics Committees specifically look at themes such as the risks and benefits of the research, confidentiality, conflict of interest and social value, all of which are also themes relevant to good data management. Not all research projects go before such an ethics committee, however, so there were further challenges in how to engage and involve researchers at the University of Warwick.

How should the policy work in practice? Refine your requirements

After the approval and publication of the Research Data Management Policy, the University of Warwick established a Research Data Management Working Group, chaired by Tim Jones, one of the PVCs for Research. This working group was established, "[…] to consider and progress initial implementation of the University's policy on Research Data Management (approved by Steering in November 2011) and related initiatives, alongside development of a profile for anticipated costs of adoption across the University."[3] The group would also undertake discussions with Monash University in Australia, which had established research data management systems and with which the University of Warwick had close contact with.

Alongside the University's Working Group, in October 2012, a Research Data Action Group was formed by the librarian, Robin Green, drawing membership from various departments including the Library, IT Services, Research Support Services, Human Resources (HR), Student Careers and Skills, and known data management experts from the Physics department and the Institute for Employment Research.

[3] Text taken from an internal report to the Senior Management Team at Warwick.

The purpose of this group was to put into action the plans made by the PVC's working group, and to draw together some of the ground-level initiatives around data management that were already happening at the University, and it was the action group that the author was particularly involved in.

Warwick's Research Data Action Group was keen to spread awareness of the University's data management policy and to find out:

- how data management affects different kinds of research (uncover good practice and existing expertise at Warwick)
- what practical support researchers would welcome
- what kind of advice and guidance would be beneficial

This Group's membership reflected the roles already being played in the arena of data management. For example, the Library's Wolfson Research Exchange[4] had already held two events on data management for researchers, and was very keen to help researchers in sharing their expertise with each other as part of the Library's mission to provide information, support and community to the University. The branch of HR responsible for supporting staff learning (the Learning and Development Centre) had also recently funded a researcher from the Politics and International Studies department, to bring in the UK Data Archive for a two day event on data management.

IT Services at Warwick had recently established an Academic Technologies Team to scope and support academics' IT needs, and Research Support Services (RSS) at Warwick are instrumental in the grant application process: the Research Data Management Policy, along with the Research Code of Practice, appears on the RSS web pages.

One of the first steps taken by this Research Data Action Group was to invite the involvement of the UK's Digital Curation Centre (DCC) in order to learn from the data management initiatives and good practice from other universities around the UK. Members of staff from the DCC were able to dedicate a number of hours work to support the University of Warwick, and the librarian co-ordinated their contribution, after introducing them to the action group.

Two workshops were held in October and December 2012, when the DCC introduced the action group to various tools that might be of use to Warwick, in particular:

- DCC's own Data Management Plan templates for the key UK funders (DCC, 2012).
- The University of Bath Roadmap, an example of good practice (University of Bath, 2012).

[4] See: http://go.warwick.ac.uk/researchexchange

- Collaborative Assessment of Research Data Infrastructure and Objectives (CARDIO), for assessing data management support and infrastructure and planning for improvement (DCC, ca.2011).
- The Data Asset Framework (DAF) that is designed to help with identifying researchers' data management activity and requirements. (University of Glasgow Human Advanced Technology & Information Institute, in association with the DCC, ca. 2012). This was once known as the Data Audit Framework, but "audit" is quite an off-putting term at UK universities, it seems.

Research your researchers: interviews

In December 2012, staff from the DCC carried out interviews with eight researchers at Warwick from the Department of Physics and the Institute of Employment Research, which were shadowed by members of the Research Data Action Group with a view to further interviews being carried out, to scope researchers' needs further.

The interviews, and indeed researchers' discussion at the data management events, revealed that researchers' focus was on how to write good data management plans when applying for research grants, and how to keep data sufficiently secure. There were differences between researchers' perceptions of what is research data. For example, the versions of code used to process a few data points might be considered data themselves, and the issues in handling data varied between the physicists and the social scientists.

Thirteen particular issues that arose from the interviews and support events attended by the author were:

1. Scepticism about the value of sharing data.
2. Keeping data might prove more costly than if it were to be reproduced at point of need.
3. Feeling threatened by a new (bureaucratic) requirement that might hamper the progress of research itself.
4. Any university practices should not interfere with existing practices for data storage and sharing in their own discipline.
5. Few researchers had already been required to submit a data management plan: an online tool to produce one and save them time would be welcome.
6. Some physicists were sceptical of the role of metadata, describing their data: this was not a priority for them.

7. In contrast, social science researchers were interested in submitting data to the UK Data Archive, but in need of time and support to get the data sets into a suitable condition to be archived and shared in this way, and this included the creation of metadata.
8. How to encrypt data and how to ensure security of data, in line with data protection requirements?
9. How to anonymise personal data?
10. How to arrange collective access amongst researchers, that is not open or public access?
11. Provide access to specialist IT equipment and support for handling and moving extremely large data sets. Moving data from one computer to the next was a big challenge for some researchers.
12. Back-up their data centrally for them: amongst interviewees, this was done by the department, or in consortium arrangements with other universities. Amongst event participants, many of whom were PhD students, some did this for themselves.
13. For some large studies, processing of data might be outsourced to a third party company: the implications of this practice need to be considered.

These thirteen issues represent Warwick's challenges going forward, after the introduction of the policy. Some can be overcome by the provision of guidelines and others by information and awareness. Others seem to require the development of practical support.

Possible solutions to these challenges

Overcoming scepticism about sharing data will likely take time and a variety of approaches, since every researcher has a different angle on the topic, much like the introduction of the institutional repository. A quote from Simon Hodson of JISC, who presented at one of the Wolfson Research Exchange's events rings true with many researchers, that "the first person to re-use your data is your future self" (Hodson, 2013). Looking at the example of the introduction of the institutional repository at Warwick, it is important that researchers are given the freedom to define good practices for their own research, sharing it with other researchers in their discipline, and thus they understand that what they are doing is truly about good research, rather than an administrative burden.

Concerns over the costs of keeping data are shared by all who are involved in data management, and researchers should be involved in the decision making process about what is kept, where and for how long: data management plans enable them to do this, in ways that meet the expectations of their own discipline and also to factor costs

associated into project plans. The University of Warwick can support its researchers in the creation of these plans and ultimately could provide a repository and/or recommendations of suitable external providers.

Researchers' anxiety over an additional burden of work calls for their involvement in setting requirements so that they are not overly onerous, and it calls for the installation of support at the university level so that the burden is reduced. Warwick's interviews were the first step in achieving this and further involvement and consultation is planned.

The importance of the different disciplinary needs and practices ought to be taken into consideration, and it is possible that the disciplinary community can already support researchers appropriately. The University should learn which researchers are not already well supported in the area of data management, and could focus on supporting their needs, whilst also understanding the practice of disciplines with advanced data management practices, in order that good practice is shared, and it is not impeded.

Some of these issues could be addressed by provision of guidelines and training opportunities, particularly through sharing expertise already in existence at the University. Some other UK universities and organisations provide such suitable guidelines already, and with support for trainers and these can be useful source material for universities like Warwick. For example, the Research Data MANTRA course made of online learning units, which is designed for PhD students and comes with a kit for librarians and trainers to use it (University of Edinburgh EDiNA, 2011). The author was working on guidelines for Warwick researchers with a colleague, at the time that she left.

Other needs represented here are for a technical infrastructure: this could be provided by the University's services or indeed by a third party company, although it is clear that the issues involving third party involvement with research data do need to be scoped further.

There are examples of technical infrastructure elsewhere that the University of Warwick could look at. The need to store and handle data for the purposes of the research itself, the storage of data for future use and making such stored data discoverable are three not incompatible needs, but might need to be kept separate. Oxford University's DataFlow system is "a two-stage data management infrastructure that is designed to allow researchers to work with, annotate, publish, and permanently store research data" (DCC, 2013), and provides an interesting model. Australia's National Data Service (ANDS) is building various tools including a Commons for descriptions of data collections and other infrastructural features that provide "connections between the data, researchers, research, instruments and institutions" which is another interesting space to watch, especially given Warwick's links with Monash University (ANDS, 2012).

Conclusion

Here we have seen, in basic steps, a university that has: defined research data, set a policy then established groups to turn the policy into action, to refine it, to research data management practices at the university and to support researchers. It has been important in these steps, to involve researchers themselves wherever possible, to use national initiatives for support and to build on the experience of other universities. The next steps for the University of Warwick will be to provide more information and training to researchers, to develop practical support, or at least to source it externally and make recommendations to researchers.

These steps are all very similar to those that Simon Hodson of JISC outlined in his visit to the University of Warwick in March 2013 (Hodson, 2013). This path begins with 'Guidance and Policy Development' and goes on to 'Training and Information'. Next comes 'Support for Data Management Planning', and then 'RDM Systems and infrastructure'. Finally, Hodson's plan features 'Publication and Citation mechanisms' (e.g. recognition, rewards and benefits for data sharing, etc), which is perhaps less well known territory, and a much bigger challenge for the University of Warwick and the academic community as a whole to get to grips with.

The University of Warwick has been able to learn from others in these steps, and this article may, in turn, provide an example for others to learn from.

References

Australian National Data Service, 2012. About ANDS: Our Approach. [online] Available at: http://www.ands.org.au/about/approach.html#ardc [Accessed 6 September 2013].

Digital Curation Centre (DCC), ca. 2011. CARDIO Collaborative Assessment of Research Data Infrastructure and Objectives. [online] Available at: http://cardio.dcc.ac.uk/ [Accessed 6 September 2013].

Digital Curation Centre (DCC), ca. 2012. DMP Online, the DCC Data Management Planning Tool. [online] Available at: https://dmponline-test.dcc.ac.uk/ [Accessed 6 September 2013].

Digital Curation Centre (DCC), 2013. Resources for digital curators: Digital curation resources from outside the DCC : DataFlow. [online] Available at: http://www.dcc.ac.uk/resources/external/dataflow [Accessed 6 September 2013].

Engineering and Physical Sciences Research Council (EPSRC), 2013. EPSRC Policy Framework on Research Data. [online] Available at: http://www.epsrc.ac.uk/about/standards/researchdata/ [Accessed 6 September 2013].

HM Government Cabinet Office, 2012. Open Data White Paper: Unleashing the potential. [online pdf] The Stationery Office. Available at: https://www.gov.uk/government/publications/open-data-white-paper-unleashing-the-potential [Accessed 2 September 2013].

Hodson, S, ca. 2013. Challenges and Opportunities in Research Data Management. [online ppt] Available at: http://www2.warwick.ac.uk/fac/cross_fac/ias/current/ace/resources2 [Accessed 6 September 2013].

Research Councils UK (RCUK), 2011. RCUK Common Principles on Data Policy. [online] Available at: http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx [Accessed 6 September 2013].

The Royal Society, 2013. Science as a Public Enterprise. [online] Available at: http://royalsociety.org/policy/projects/science-public-enterprise/ [Accessed 22 August 2013].

UK Data Archive, 2013. UK Data Archive - ABOUT US. [online] Available at: http://data-

archive.ac.uk/about [Accessed 2 September 2013].

University of Bath, 2012. University of Bath Roadmap for EPSRC Compliance with Research Data Management Expectations. [online] Available at: http://www.bath.ac.uk/rdso/University-of-Bath-Roadmap-for-EPSRC.pdf [Accessed 6 September 2013].

University of Edinburgh EDiNA, ca. 2011. MANTRA, Research Data Management Training. [online] Available at: http://datalib.edina.ac.uk/mantra/ [Accessed 6 September 2013].

University of Glasgow Human Advanced Technology & Information Institute, in association with the DCC, ca. 2012. Data Asset Framework.[online] Available at: <http://www.data-audit.eu/index.html> [Accessed 6 September 2013].

University of Warwick, 2011. Research Data Management Policy. [online] Available at: http://www2.warwick.ac.uk/services/rss/researchgovernance_ethics/research_code_of_practice/datacollection_retention/reseatch_data_mgt_policy [Accessed 6 September 2013].

University of Warwick, 2012. Research Ethics Committees (RECs). [online] Available at: http://www2.warwick.ac.uk/services/rss/researchgovernance_ethics/research_code_of_practice/researchethicscommittees [Accessed 6 September 2013].

University of Warwick, 2013. University of Warwick Publications service & WRAP: Browse by Year. [online] Available at: http://wrap.warwick.ac.uk/view/publications_year/ [Accessed 6 September 2013].