cilip

Chartered Institute of
Library and Information
Professionals

Research Data Management Briefing Paper (July 2014)

This briefing paper defines Research Data Management (RDM) and sets out the drivers for better RDM. It briefly considers how academic libraries support research and the opportunities and challenges RDM presents for library and information professionals. A brief overview of current RDM activity within libraries is included.

Contents

1. Introduction

"Technology has enabled data to become the prevalent material and currency of research. Data, not information, not publications, is rapidly becoming the accepted deliverable of research." - Graham Pryor, 2012. *Re-skilling for research - observations on an RLUK report*

The output from research in the 21st century is predominantly data, produced mainly in electronic form. This data is being generated at a rapid rate; a phenomenon often referred to as the "data deluge".[1]

If the information contained in digital materials is to remain retrievable, identifiable and usable, it needs to be managed from the outset, because digital materials are vulnerable to storage failures and technological obsolescence.[2] At present, only a small

---

[1] Pryor, G. Why Manage Research Data? In Pryor, G. (Ed), 2012, *Managing Research Data*, p2
[2] Cox, A. and Pinfield, S., 2013. Research data management and libraries: Current activities and future priorities, *Journal of Librarianship and Information Science*, Vol 45 (4)

proportion of data is correctly managed and made accessible for use and reuse. However, with the growth of RDM, this is changing.

2. Definitions

In academic research, data is "the output from any systematic investigation involving a process of observation, experiment or the testing of a hypothesis, which when assembled in context and interpreted expertly will produce new knowledge".[3]

Research data includes analogue sources, as well as "discrete digital objects" (e.g. text files, image files or sound files), "complex digital objects" (discrete digital objects made by combining a number of other digital objects, such as websites) and databases.[4]

RDM is "the organisation of data, from its entry to the research lifecycle through to the dissemination and archiving of valuable results" (Whyte and Tedds, 2011).[5] It is undertaken to ensure the long-term value of the data and consists of a number of different activities and processes associated with the data lifecycle, which must take account of technical capabilities, ethical considerations, legal issues and governance frameworks.[6] When associated with digital data, these activities and processes are known collectively as digital curation.

3. The digital curation lifecycle

The Digital Curation Centre is a centre of expertise in digital information curation with a focus on building capacity, capability and skills for research data management across the UK's higher education research community.[7] It has produced a Curation Lifecycle Model to provide a graphical, high-level overview of the stages required for successful curation and preservation of data. The model can be downloaded from the DCC website: http://www.dcc.ac.uk/resources/curation-lifecycle-model

The digital curation lifecycle comprises of the following actions[8]:

Conceptualise: conceive and plan the creation of digital objects, including data capture methods and storage options.

*Create***:** produce digital objects and assign metadata.

Access and use: ensure that designated users can easily access digital objects (a password may be required).

---

[3] Pryor, G. Why Manage Research Data? In Pryor, G. (Ed), 2012, *Managing Research Data*, p2
[4] Ibid, p7
[5] Whyte, A and Tedds, J., 2011. *Making the Case for Research Data Management*, Digital Curation Centre
[6] Cox, A. and Pinfield, S., 2013. Research data management and libraries: Current activities and future priorities, *Journal of Librarianship and Information Science*, Vol 45 (4)
[7] http://www.dcc.ac.uk/about-us
[8] See http://www.dcc.ac.uk/resources/curation-lifecycle-model

Appraise and select: evaluate digital objects and select those requiring long-term curation and preservation. Adhere to documented guidance, policies and legal requirements.

Dispose: rid systems of digital objects not selected for long-term curation and preservation. Once again adhere to documented guidance, policies and legal requirements.

Ingest: transfer digital objects to an archive, trusted digital repository, data centre or similar, again adhering to documented guidance, policies and legal requirements.

Preservation action: undertake actions to ensure the long-term preservation and retention of the authoritative nature of digital objects.

Reappraise: return digital objects that fail validation procedures for further appraisal and reselection.

Store: keep the data in a secure manner as outlined by relevant standards.

Access and reuse: ensure that data are accessible to designated users for first time use and reuse. Some material may be publicly available, whilst other data may be password protected.

Transform: create new digital objects from the original, for example, by format shifting.

4. What are the drivers for RDM?

> "The availability of research data is high on the agenda of higher education policy makers, funders and researchers committed to open practice" - Simon Hodson and Sarah Jones, 2013, *Seven rules of successful research data management in universities*, Guardian

Some of the drivers for RDM are discussed below:

4.1 Data as a public good

Every year £3.5b of taxpayers' money is spent on research undertaken by UK universities.[9] There is a mounting expectation that the results of this research should be made as widely available as possible for the public good.   This has been a driver behind the "open access" movement, a worldwide effort to provide free online access to

---

[9] Pryor, G. Why Manage Research Data? In Pryor, G. (Ed), 2012, *Managing Research Data*, p1

scholarly research literature.[10] Many of the arguments for open access to publically funded research literature apply equally well to publically funded research data.

Data may often have uses unforeseen by the original creators and further information may be extracted by applying different techniques or integrating with other data sets.[11] For example, the Royal Society's report, Science as an Open Enterprise[12], stresses the potential for data reuse and a need for rapid data sharing to respond to global challenges, such as flu epidemics. However, at present there is only patchy coverage of subject-specific repositories.[13] Without RDM, the data deluge is likely to result in lost opportunities and considerable financial waste.

4.2 *Funders' requirements*

> "The expectation by the major funders that research data as a recognised asset will be afforded due care and attention has become more overt in recent years, confirmed by an emerging requirement for the inclusion of data management plans within research grant proposals".- Graham Pryor, 2012, *Managing Research Data*, p4

Many funding bodies, including all the UK Research Councils, now mandate that applicants produce a RDM plan and that data sharing and reuse is designed into it where possible.

To foster good practice, Research Councils UK (RCUK) has coordinated a statement of Common Principles on Data Policy asserting that "making research data available to users is a core part of the Research Councils' remit". The statement contains seven common principles of data policy, with the intention of providing an overarching framework for individual research council policies. It is available at hhttp://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx

The DCC tracks and summarises funder policies. "An Overview of Funders' Data Policies" is available from:
http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies

---

[10] See CILIP's Open Access briefing paper
http://www.cilip.org.uk/sites/default/files/documents/CILIP%20briefing%20on%20the%20open%20access%20movement%20July%202013.pdf
[11] Hodson, S. and Jones, S., 2013 Seven rules of successful research data management in universities, Guardian http://www.theguardian.com/higher-education-network/blog/2013/jul/16/research-data-management-top-tips
[12] http://royalsociety.org/policy/projects/science-public-enterprise/report/
[13] Simmonds et al., 2011 cited in Cox, A. and Pinfield, S., 2013. Research Data Management and Libraries: Current Activities and Future Priorities. *Journal of Librarianship and Information Science*, Vol 45 (4), p1

The Engineering and Physical Sciences Research Council (EPSRC), for example, now requires research organisations to preserve data securely for at least 10 years, and "… ensure that effective data curation is provided throughout the full data lifecycle".[14]

4.2 Research integrity

RDM is a means to ensure research integrity. It can also mitigate risks to institutional reputation.

Sound research rests on the ability to evidence, verify and reproduce results. Two recent cases that illustrate the need for good RDM are "Climategate" and errors in the Reinhart-Rogoff paper.

Climategate

In November 2009, over 1,000 private emails between climate change scientists from the University of East Anglia's (UEA) Climatic Research Unit (CRU) were stolen or leaked and published online. Selected contents of these emails were used by some to suggest that scientists had been manipulating or hiding climate change data and had deliberately frustrated requests for access to the data to prevent the publication of papers they disagreed with. Climate sceptics believed they had found the "smoking gun" which confirmed their belief that global warming was not happening.[15]   This controversy was dubbed "Climategate".

It emerged that the CRU had been adjusting data because the original data had been disposed of. Without the original data other scientists could not verify the accuracy of the "adjusted" data nor the conclusions drawn from it. A House of Commons Science and Technology Select Committee inquiry concluded that the reputation of the CRU was intact, but there were issues over the management of its data.[16] Likewise the Independent Science Assessment Panel concluded that there was no evidence of any deliberate malpractice in the work of the CRU scientists, but suggested that the Unit could have done more to document and archive its data.[17]

Reinhart-Rogoff paper on economic growth

Politicians, commentators and activists widely cited "Growth in a time of debt", a paper by economists Carmen Reinhart and Kenneth Rogoff, published in the American Economic Review in 2010, in political debates over the effectiveness of austerity in fiscal policy for debt-burdened economies. However the paper was subsequently found to contain errors and the exclusion of some data that significantly undermined the

---

[14] http://www.epsrc.ac.uk/Pages/default.aspx
[15] http://www.theguardian.com/environment/2010/jul/07/climate-emails-question-answer
[16] http://www.publications.parliament.uk/pa/cm201011/cmselect/cmsctech/444/444.pdf
[17] http://www.uea.ac.uk/mac/comm/media/press/CRUstatements/SAP

results. The American Economic Review had failed to enforce its own data availability policy, which meant that these errors were only discovered in 2013.[18]

The UK Research Integrity Office (UKRIO) states in its Code of Practice:

"Organisations should have in place procedures, resources (including physical space) and administrative support to assist researchers in the accurate and efficient collection of data and its storage in a secure and accessible form. Researchers should consider how data will be gathered, analysed and managed, and how and in what form relevant data will eventually be made available to others, at an early stage of the design of the project."[19]

4.3 Regulatory requirements and intellectual property

Measures to comply with Freedom of Information and Data Protection legislation need constant monitoring. Copyright needs consideration too.

Freedom of Information

Freedom of Information gives everyone the right to request information held by public sector organisations under the Freedom of Information Act 2000. Research data can be the subject of Freedom of Information requests, as was the case in Climategate. Partly in response to Climategate, in 2010 JISC developed new guidance for researchers in responding to FOI requests for research data.[20]

Data Protection

Research data about individuals may be subject to the Data Protection Act 1998. The Act has eight principles relating to the way data can be obtained, held, used or disclosed. These include   the principles of security, accuracy and relevancy.

The EU is in the process of reforming data protection policy and issuing a Data Protection Regulation with direct implementation in EU countries.   Final agreement is expected during 2015, with enforcement of the Regulation beginning in 2017.[21] This is likely to have implications for effective RDM.

---

[18] http://www.theguardian.com/higher-education-network/blog/2013/jul/16/research-data-management-top-tips
[19] UK Research Integrity Office 'Code of Practice for Research', 2009. (p.17) Available from: http://www.ukrio.org/what-we-do/code-of-practice-for-research/
[20] JISC, 2010. *New advice for universities in light of the Climate Change Emails Review*. Available from http://www.jisc.ac.uk/news/stories/2010/07/opendata.aspx
[21] http://www.eudataprotectionlaw.com/timeline/

Intellectual Property

Intellectual property rights include copyright, patents, trademarks and design rights. These rights should be identified and ownership clarified at the beginning of the research process to avoid unexpected restrictions on how data may be used and reused.

5. Universities' response to RDM

In response to these drivers the management of research data has emerged as a strategic priority for universities. Some UK universities have started to develop RDM policies and an overview is available from DCC at http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies/uk-institutional-data-policies

The policies do the following:[22]

- Identify areas of responsibility for the institution and for researchers
- Commit the university to develop appropriate guidelines, training and support, including mechanisms and services for storage and backup
- Support deployment of data repositories and/ or mechanisms for registering metadata about research data
- Recognise that management and curation of research data requires cooperation and coordination with research funders, and with existing national and international providers of data services and subject-based repositories

These policies build on earlier work supported by the JISC Digital Repositories and Preservation programme in the projects "Embedding Institutional Data Curation Services in Research" (EIDCSR) and DataShare respectively[23]. JISC considers it a priority to support universities in improving RDM.

6. Libraries and support for research

In 2011 Research Information Network (RIN) and Research Libraries UK (RLUK) jointly commissioned a study into the value of the services that libraries in the UK provide to researchers, and of the contributions that libraries make to institutional research performance.[24] The report found that:

---

[22] Taken from Whyte, A. and Tedds, J., 2011. *Making the Case for Research Data Management* http://www.dcc.ac.uk/resources/briefing-papers/making-case-rdm#
[23] Ibid
[24] Research Libraries UK & Research Information Network, 2011. The *value of libraries for research and researchers* http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/value-libraries-research-and-researchers

- Good libraries help institutions to recruit and retain top researchers
- Good libraries help researchers win research grants and contracts
- Repositories, which in most cases are run by the library, increase the visibility of the institution and raise its research profile
- Researchers who make use of specialist library staff see them as vital, but too often specialists and researchers are not well connected. Connecting with researchers enhances the value of the library's services. Many researchers do not use the physical library and, as remote online access is becoming quick, seamless and easy, libraries are becoming invisible. Reconnecting with researchers and understanding their needs can increase "researcher satisfaction, as well as winning recognition and respect for the library across the institution" (p47).

The report concluded that the need for academic libraries "to demonstrate value will endure and should not be underestimated. Arguing the case for libraries may get harder as the traditional roles of libraries in providing access to content… continues to become less visible….A big challenge for libraries is to communicate to both researchers and senior managers how they are changing, and the opportunities for the future." (p64)

7. Does RDM present an opportunity for library and information professionals?

RDM "is a concept with limited appeal to the majority of a research community that receives short-term funding and is composed of a highly mobile workforce... the terminology used in the DCC curation lifecycle model also suggests a different skill set to that traditionally associated with researchers, one that instead implies the stewardship and husbandry of data rather than its active use… The greatest resonance with this model is with the information practitioner" – Graham Pryor, 2012. *Managing Research Data*, p9

The growth of RDM potentially creates a number of specialist roles and it has been suggested that academic library services are well positioned to take on all or some of these. For example, generic information management skills could be extended to RDM and the provision of RDM advice and training could be regarded as an extension of existing library work in advice services and information literacy training.[25]

Cox et al (2012) have listed the main roles that have been proposed, and points to links in existing library practices and to the corresponding professional knowledge base. This is set out in a table that is available at:
http://www.ariadne.ac.uk/issue70/cox-et-al#3[26]

---

[25] Cox, A. and Pinfield, S., 2013. Research Data Management and Libraries: Current Activities and Future Priorities. *Journal of Librarianship and Information Science*, Vol 45 (4)

[26] Librarians' roles in RDM and required competencies mapped to existing roles. Taken from Cox, A., Verbaan, E., and Sen, B., 2012. Upskilling Liaison Librarians for Research Data Management. *ARIADNE*, issue 70 http://www.ariadne.ac.uk/issue70/cox-et-al#3. For an adapted version see Cox, A. and Pinfield,

LIBER, the Association of European Research Libraries, has made 10 recommendations for libraries to get started with RDM:[27]

- Offer support, including IP advice and funding advice
- Engage in development of metadata and data standards
- Create data librarian posts
- Actively participate in institutional research data policy development. Encourage and adopt open data policies
- Liaise and partner with researchers, data archives and centres to foster an interoperable infrastructure for access, discovery and sharing
- Provide services for storage, discovery and permanent access
- Promote research data citation by applying persistent identifiers
- Provide a catalogue or repository
- Get involved in subject specific data management
- Offer or mediate secure storage for dynamic and static research data in cooperation with IT or seek exploitation of cloud services

8. Challenges presented by RDM

Despite these opportunities, there are challenges for libraries and library and information professionals to overcome before they can fully engage with RDM:

- Translating current library expertise and practices, such as in metadata creation, to research data contexts is not straightforward[28]. RDM is jargon ridden and the technical knowledge required for some roles is significant. It is more closely tied to records management and archival thinking than librarianship[29].
- Librarians often lack direct personal experience of research[30]
- There are problems engaging users, especially researchers, with LIS services[31]
- Academic librarians are already very busy. Supporting RDM may mean down-rating other priorities.[32]
- Resources, infrastructure, policy and governance structures are still in flux.[33]

S., 2013. Research Data Management and Libraries: Current Activities and Future Priorities. *Journal of Librarianship and Information Science*, Vol 45 (4), p4

[27] LIBER, 2012. Ten recommendations for libraries to get started with research data management. Available at http://www.libereurope.eu/sites/default/files/The%20research%20data%20group%202012%20v7%20final .pdf

[28] Librarians' roles in RDM and required competencies mapped to existing roles. Taken from Cox, A., Verbaan, E., and Sen, B., 2012. Upskilling Liaison Librarians for Research Data Management. *ARIADNE*, issue 70

[29] Cox, A. and Pinfield, S., 2013. Research Data Management and Libraries: Current Activities and Future Priorities. *Journal of Librarianship and Information Science*, Vol 45 (4)

[30] Librarians' roles in RDM and required competencies mapped to existing roles. Taken from Cox, A., Verbaan, E., and Sen, B., 2012. Upskilling Liaison Librarians for Research Data Management. *ARIADNE*, issue 70

[31] Ibid

[32] Ibid

[33] Ibid

- The provision of professional education and training in digital curation is far more extensive in the USA that in the UK.

In an investigation into the skills needed by subject and liaison librarians to support the evolving information needs of researchers, Mary Auckland found a potentially significant gap in the knowledge required to advise on RDM, even though 48% of respondents to her survey, which was conducted in 2012, felt that, within 2-5 years, this knowledge would be considered essential.[34]

9. Current RDM activity within academic libraries

In 2012 Cox and Pinfield surveyed UK universities to understand the way libraries are currently involved in RDM. Although their research concluded that, at the end of 2012, academic libraries offered only limited RDM services, the evidence suggested that this is changing. In the short term, developing RDM policy was found to be the priority, with 31% of responding institutions already having a policy in place, and 43% expected to have one by the end of 2013. In the medium term, the focus is likely to shift to advice and training, as well as some involvement in running a data repository.[35]

RDM isn't named as a concept in the RIN & RLUK report referred to above, but a number of services integral to RDM feature in it.[36] For example, some libraries were found to be supporting research efforts across their institutions by providing services such as:

- Managing the open access repository/research archive
- Advising researchers on topics (e.g. open access, data security) to help the institution meet compliance requirements of the research and funding councils
- Facilitating data sharing between the open access repository and other systems
- Assisting with research grant proposals
- Supplying respected knowledge and a valued contribution to debates and current issues

10. Further reading, case studies and resources

- Chad, K. and Enright, S., 2014. The Research Cycle and Research Data Management (RDM): Innovating Approaches at the University of Westminster. *Insights*, Vol 27 (2), pp147-153

---

[34] Auckland, M., 2012. *Re-skilling for Research*, Research Libraries UK
http://www.rluk.ac.uk/files/RLUK%20Re-skilling.pdf
[35] Cox, A. and Pinfield, S., 2013. Research Data Management and Libraries: Current Activities and Future Priorities. *Journal of Librarianship and Information Science*, Vol 45 (4)
[36] Research Libraries UK & Research Information Network, 2011. The *value of libraries for research and researchers* http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/value-libraries-research-and-researchers

- Cox, A. and Pinfield, S., 2013. Research Data Management and Libraries: Current Activities and Future Priorities. *Journal of Librarianship and Information Science*, Vol 45 (4)
- Cox, A., Verbaan, E., and Sen, B., 2012. Upskilling Liaison Librarians for Research Data Management. *ARIADNE*, issue 70 http://www.ariadne.ac.uk/issue70/cox-et-al#3
- Hodson, S. and Jones, S., 2013 *Seven rules of successful research data management in universities*, Guardian http://www.theguardian.com/higher-education-network/blog/2013/jul/16/research-data-management-top-tips
- Pickton, M., Jones, S., and Guy, M., 2012*.* Leading from the library: data management initiatives at the University of Northampton. Paper presented to: The International Federation of Library Associations and Institutions (IFLA) World Library and Information Congress: 78th IFLA General Conference and Assembly (IFLA2012), Helsinki, Finland, 11-17 August 2012*.* http://nectar.northampton.ac.uk/4425/ [accessed 22 July 2014]
- Pryor, G. (Ed), 2012. *Managing Research Data*. London: Facet Publishing
- Pryor, G., Jones, S., and Whyte, A., 2014. *Delivering Research Data Management Services. London*: Facet Publishing
- Whyte, A and Tedds, J., 2011. *Making the Case for Research Data Management*, Digital Curation Centre: http://www.dcc.ac.uk/resources/briefing-papers/making-case-rdm#Introduction
- RDMRose was a JISC funded project to produce taught and continuing professional development (CPD) learning materials in Research Data Management (RDM) tailored for Information professionals: http://www.sheffield.ac.uk/is/research/projects/rdmrose
- DCC's digital curation training for all: http://www.dcc.ac.uk/training
- Discipline guides and case studies from the University of Hertfordshire: http://www.herts.ac.uk/rdm/training/discipline-guidance
- Presentations from CILIP's Research data Management Executive briefing (May 20th 2014): http://www.cilip.org.uk/cilip/events/cilip-research-data-management-executive-briefing