

■ SUPPORTING RESEARCH DATA MANAGEMENT AND OPEN SCIENCE IN ACADEMIC LIBRARIES: A DATA LIBRARIAN'S VIEW

by *Robin Rice*

Abstract: *The ‘data revolution’ has impacted researchers across the disciplines. As if the traditional work of teaching, competing for grants and promotion, doing research and publishing results was not challenging enough, researchers are required to make fundamental changes in the way they do all of these things. A similar shift can be seen for academic librarians. Librarians who were taught to meet the needs of their users based on information scarcity now need to retrain themselves to help users deal with information overload. Moreover, librarians increasingly find themselves ‘upstream’ in the research process, trying to assist their users in managing unwieldy amounts of data when their comfort zone is firmly ‘downstream’ in the post-publication stage. Unsettling as it may be, these are exciting developments for the library profession.*

Keywords: *Research data management; Open Science; Academic Libraries*

UNTERSTÜTZUNG VON FORSCHUNGSDATENMANAGEMENT UND OFFENER WISSENSCHAFT IN WISSENSCHAFTLICHEN BIBLIOTHEKEN: DIE SICHT EINES DATA LIBRARIANS

Zusammenfassung: *Die „Datenrevolution“ hat sich auf die WissenschaftlerInnen aller Disziplinen ausgewirkt. Als ob die traditionelle Arbeit in der Lehre, im Wettbewerb um Stipendien und Förderungen, in der Forschung und bei der Veröffentlichung von Forschungsergebnissen nicht schon anspruchsvoll genug wäre, müssen sie grundlegende Veränderungen in der Art und Weise vornehmen, wie sie all diese Dinge tun. Eine ähnliche Veränderung ist bei den wissenschaftlichen BibliothekarInnen zu beobachten. BibliothekarInnen, denen beigebracht wurde, die Bedürfnisse ihrer NutzerInnen aufgrund von Informationsknappheit zu befriedigen, müssen nun umlernen, um den NutzerInnen bei der Bewältigung der Informationsflut zu helfen. Darüber hinaus finden sich BibliothekarInnen bereits zunehmend im Forschungsprozess wieder und versuchen, ihre NutzerInnen bei der Verwaltung sperriger Datenmengen zu unterstützen. So verunsichernd dies auch sein mag, sind dies spannende Entwicklungen für den Bibliotheksberuf.*

Schlagwörter: *Forschungsdatenmanagement; Open Science; wissenschaftliche Bibliotheken*

DOI: <https://doi.org/10.31263/voebm.v72i2.3303>



Parts of this paper were first published as part of the Indian Institute for Management, Bangalore Future of Libraries conference proceedings.

1. Introduction

The ‘data revolution’ has impacted researchers across the disciplines. As if the traditional work of teaching, competing for grants and promotion, doing research and publishing results was not challenging enough, researchers are required to make fundamental changes in the way they do all of these things: teaching must make use of learner ‘analytics’; bureaucratic decisions must be ‘data-driven’; research must be digital, with transparent methods, making use of ‘big data’; publishing results must be open access, and data sharing is often a prerequisite for a paper to be published.

A similar shift can be seen for academic librarians. Librarians who were taught to meet the needs of their users based on information scarcity now need to retrain themselves to help users deal with information overload. Moreover, librarians increasingly find themselves ‘upstream’ in the research process, trying to assist their users in managing unwieldy amounts of data when their comfort zone is firmly ‘downstream’ in the post-publication stage.

Unsettling as it may be, these are exciting developments for the library profession. Researchers who long stopped using reference services to conduct their own searches over the Internet are seeking professional help for their research data management (RDM), and are increasingly finding that expert help from librarians. In this sense, every librarian is becoming a ‘data librarian’ in some ways.

2. The Inside Out Library

A 2016 radical experiment resulted from a University-wide task force about the future of the Library at MIT: the entire collections budget was placed under the management of the scholarly communications department. The Library management made this decision as “part of a broader strategic pivot in which research libraries focus more on “inside out” collections – those in fewer collections, often generated by the university, often unique

to that university – and less on “outside in” collections – those we buy from external sources to make available locally, and which appear in many universities’ collections.”¹

Informed by theory provided by Lorcan Dempsey of OCLC, the aim is to target financial and other resources towards collection types that have high ‘uniqueness’ value: including those that are already currently highly stewarded – such as manuscripts and other special collections, but also new information types like research data, which are not currently highly stewarded in libraries.

3. The University of Edinburgh Research Data Policy and Services

Although the University of Edinburgh has not taken radical steps with its collection budget like MIT, it serves as an example of a well-invested service focusing on University members’ research data.

Edinburgh’s research data management (RDM) programme began before 2011, when its RDM policy came into force, with a focus on requiring data management plans of every new research project (<https://www.ed.ac.uk/is/research-data-policy>). This helps both to ensure adequate provisions are made for RDM requirements in terms of costs, and that data which can be openly shared are identified early and documented well.

The policy outlines the researcher’s own responsibilities when it comes to research integrity and managing research well as part of that, and also the role of the institution in supplying tools and support in RDM across the data lifecycle, from the creation stage, to the active storage and analysis stage, to the archiving and sharing stage. At the same time, training and awareness raises the visibility of both the need for RDM and the services available to the researchers. This is communicated to researchers through a research lifecycle diagram (Figure 1) in order for them to find the right tool at the right time of their project, or ‘user journey’ (<http://www.digit-alresearchservices.ed.ac.uk/>).

While training is not mandatory, the team strives to develop rich relationships with schools (academic departments, institutes and centres) so that a meeting or event occurs with each one at least once through the academic year. The team’s communication plan makes this a concrete objective, which can be measured; at this point, it is still aspirational. The team also works with the Institute of Academic Development, Academic Service Librarians and Digital Skills, as well as departmental support staff to embed RDM training where training is sought.

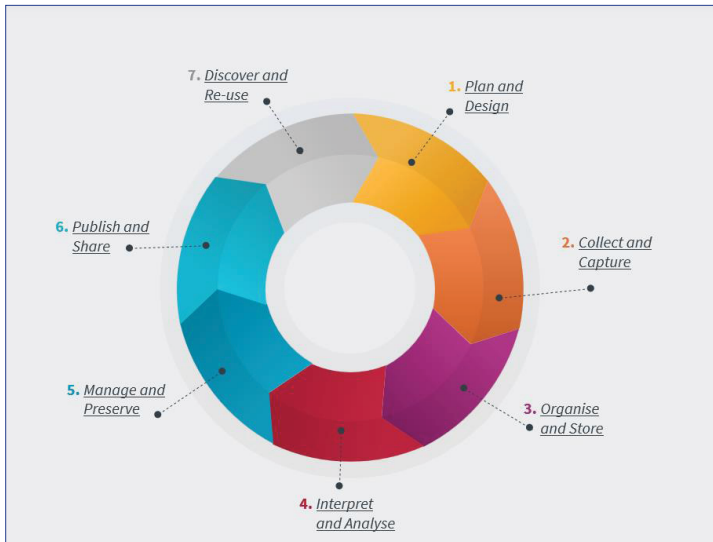


Fig. 1: University of Edinburgh Digital Research Services lifecycle approach

4. Awareness-raising, data training and support

Increasingly libraries are becoming the locus of support for research data management (RDM) enquiries. Librarians are well-placed to offer advice on local and external RDM tools and services, best practice and funder or publisher requirements. Forms of support may include a dedicated RDM website; expert advice (often mediated through a general helpdesk system); assistance with writing data management plans (DMPs); templates and examples for DMPs; pointers to general RDM training resources or face to face training options; hands-on data-related training in the use of popular tools (this can include Data Carpentry and Software Carpentry workshops); promotion and advocacy for RDM and Open Science in the form of blog posts, outreach, updates to staff meetings; and hosting or co-hosting topical or networking events for researchers.

5. Data Management Planning support

The importance of data management planning – in addition to complying with funders’ requirements – becomes apparent in two scenarios. First,

projects that intend to collect large amounts of data ('big data'), or data with large file sizes (such as medical imaging), must resource RDM sufficiently – including short and long-term storage costs and including a staff role for data management within the project. Similarly, costs for high performance computing and software must be included, though this is not normally considered part of RDM. Researchers often believe their proposals will be penalised for adding RDM costs, but increasingly the opposite is true – a well-costed grant proposal is expected to include RDM costs by mainstream funders, who understand it cannot be done well for free.

Second, projects collecting personal or sensitive data must ensure they plan for adequately safeguarding the data during the life of the project (which may entail more expensive storage environments or more rigorous organisational procedures which need to be monitored), and also for what will happen to the data when the project comes to a close – whether it needs to be destroyed or retained, for how long, and who may have access to the data, or whether an anonymised version of the data can be created and openly shared. Many of the horror stories about data loss or data breaches, as well as concerns about data sharing, can be avoided through proper data management planning, hence the policy emphasis on early creation of a DMP, whether or not the funder requires one.

6. Support for data sharing and curation

On approaching completion, researchers generally need to select an archive for data that should be kept or shared after the end of the project. This may be a national or international data archive associated with their disciplinary area (such as DANS in the Netherlands, or Dryad for biology and ecology). Universities may accept datasets into their institutional repository. In such cases, attention to persistent identifiers (such as DataCite DOIs), metadata and digital preservation is required.

As part of advocacy around open science or open research, it is recommended that researchers link their datasets with other research outputs such as articles, and their own ORCID author identifier.

7. A maturity model for RDM services

Librarians wishing to offer RDM services need to build up their expertise and offerings gradually, ensuring their efforts are matched by the expect-

tations and requirements of their users. Andrew Cox and the co-authors of the maturity model pictured in Figure 2 demonstrate how a library can begin at ‘level 0’ by conducting needs assessments using tools such as data audits and user surveys, while simultaneously developing expertise in staff through professional development activities. This then leads to a ‘level 1’ maturity characterised by minimal compliance with funders’ requirements and development of an institutional policy.

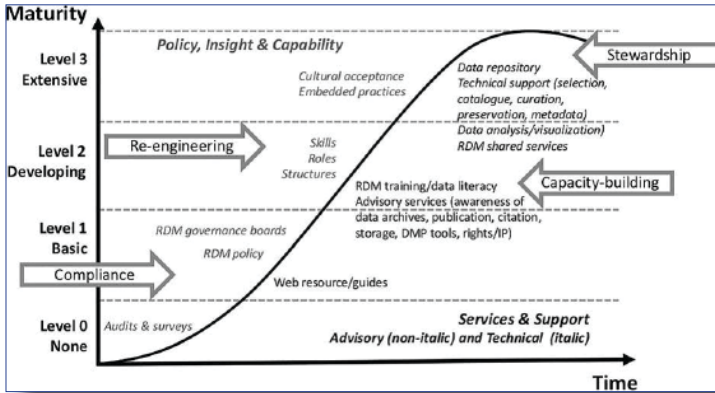


Fig. 2: A maturity model for RDM services²

While the maturity of the service is in the development stage, a re-engineering of library structures and staffing occurs to place additional resources into building librarians’ skills and changing their roles to be able to meet increased demand. The effect on the user community is that of capacity-building, with RDM and data literacy training being offered and the advisory services becoming more sophisticated and responsive. In the highest “extensive” level of the maturity model, the service is characterised by “policy, insight and capability,” wherein the user community experiences cultural acceptance of changed, embedded practices in RDM. The library will focus its energy on data stewardship, typically building a data repository service with associated technical support, and potentially be involved in a wider community of providers through participation in shared services.

8. RDM and the Open Science/Research Agenda

While a research data service may mature over time, it is also the case that the wider research environment changes over time, and the service

must adapt its key messages and activities to respond. The Open Science or Open Research agenda has now inarguably become a key driver for good practices in research data management, with an emphasis on data sharing. But Open Science encompasses a whole range of researcher behaviours, which in order to realise the full value in publicly funded research and generate improved public trust in science and scholarship, is seen as needing to change. The European Union funded project, FOSTER Open Science, explains that: *Open Science is frequently defined as an umbrella term that involves various movements aiming to remove the barriers for sharing any kind of output, resources, methods or tools, at any stage of the research process. As such, open access to publications, open research data, open source software, open collaboration, open peer review, open notebooks, open educational resources, open monographs, citizen science, or research crowdfunding, fall into the boundaries of Open Science. Even though, especially for the library and information domain, the focus is usually placed on two of these movements: Open Research Data and Open Access to scientific publications.*³

9. Open data and FAIR data

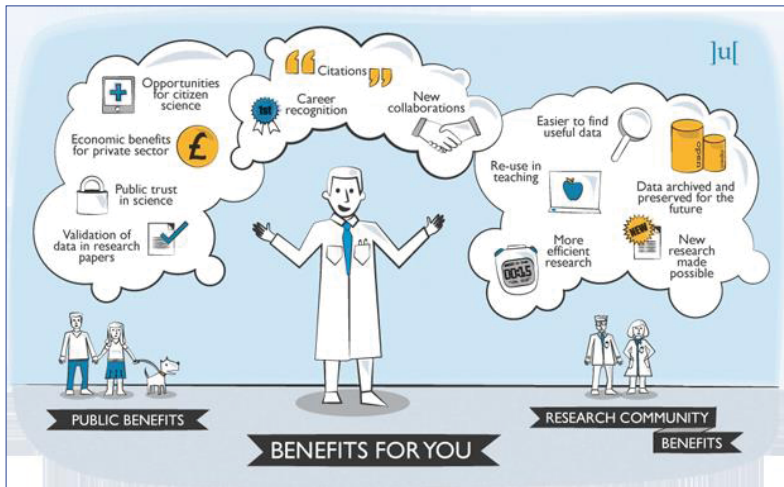


Fig. 3: Benefits of Open Data (Journal of Open Archaeology Data, CC-BY licence)

The benefits of sharing the underlying data from a research project, especially those underlying published research results, are many (see Figure 3). However individual researchers may not accept that making

their data publicly available is beneficial to them. In some senses, the benefits accrue to others – funders who do not need to pay for repeat studies; other researchers who can make do with secondary data instead of spending time on collecting their own; citizen scientists; data journalists; even, in some cases commercial companies. This is why it is crucial for librarians to emphasise the benefits of data sharing and not only compliance factors. For example, there is scattered evidence that papers, where data are shared, are more highly cited.⁴ However, as the Open Science movement grows, the distorted academic reward system based on publish versus perish and false metrics such as journal impact factors are meant to be overcome by other values, so for example, researchers can be rewarded for highly cited datasets on their own merit, not only for formal publications.

But while open data may still have its detractors, a newer concept has emerged which is even more difficult to find fault with: make your data FAIR (findable, accessible, interoperable and reusable). According to the EU-funded GoFAIR project, the acronym can be explained as follows:

FINDABLE: “Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services.”

ACCESSIBLE: “Once the user finds the required data, she/he needs to know how [they can] be accessed, possibly including authentication and authorisation.”

INTEROPERABLE: “The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.”

REUSABLE: “The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.”⁵

Among other things, this means that data that are not appropriate to be openly shared, such as personal and sensitive data, do not have to be, or due to legislative requirements such as GDPR must not be. However, the metadata describing the research data can and should be open and discoverable, and the instructions for requesting access should be clear – and preferably even machine-actionable, with full documentation made available in order to be able to reuse the data when a request is approved. For example, the European Commission has described its data sharing policy for Horizon 2020 funded research projects as “open by default”, or “as open as possible, as closed as necessary.”⁶

10. How do librarians need to reskill for data support?

There is some concern that library schools do not know how to prepare the librarians for the future of Open Science and Scholarship, and equip them with the digital and research data skills that library users of the future will expect of them. According to a meta-analysis examining results of similar surveys in four countries (Australia, Ireland, Netherlands and United Kingdom), it is indeed both data curation skills as well as data description and documentation abilities that score high, along with legal, policy and advisory skills, as well as knowledge of a variety of research methods (see Figure 4).

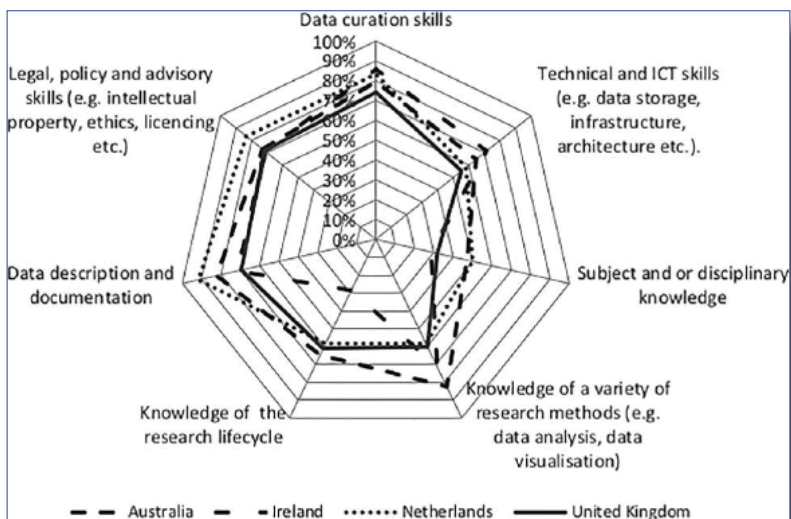


Fig. 4: Changing skills and priorities in academic libraries? (A. Cox et al.)

This might seem to point to the need for some academic librarians to specialise in data skills, or become dedicated data librarians. Yet in another study by Lisa Federer, librarians who do data-related work were surveyed about their work and educational backgrounds and asked to rate the relevance of a set of data-related skills and knowledge to their work.⁷ The surprise here was that this although traditional library skills were not rated highly by this group, ‘soft skills’ were rated very highly:

- “Personal Attributes” were the most highly rated category overall (70% respondents ranked Very important +).
- “Library Skills” were the lowest rated category (40%). Perhaps these data experts find that there is simply too wide of a gap between how

librarians are typically using their skills (e.g. reference work and cataloguing) and how these types of skills can be used in data support. It could be interesting to see if this perception changes over time, if librarians do learn skills needed for data support from the library schools of the future.

- The top five rated elements were “Developing relationships with researchers, faculty, etc.”; “Oral communication and presentation skills”; “Teamwork and interpersonal skills”; “Written communication skills”; and “One-on-one consultation or instruction.”
- The bottom five rated elements were “PhD or doctoral degree”; “Professional memberships”; “Cataloging”; “Graduate degree in a [subject discipline]”; & “Collection development.”

11. Getting started in RDM support

In addition to the learning resources mentioned in this paper, there are some excellent starting points for academic librarians wishing to make a start in the provision of research data services. These include:

1. A **top ten list** of recommendations for libraries to get started with research data management from LIBER (<https://bit.ly/2NuUhAs>)
2. Research Data Alliance (RDA) **23 things** (<http://bit.ly/RDAthing1>)
3. **LEARN RDM Toolkit** including a model policy (<https://bit.ly/2oaL0nN>)

Finally, both the League of European Research Universities (LERU)⁸ and the Association of European Research Libraries (LIBER)⁹ produced roadmaps in 2018 with accompanying tools such as checklists for institutions to prepare themselves for Open Science. As RDM becomes increasingly embedded in general Open Science strategies and technologies, it is important that a holistic approach to research support for RDM and FAIR data is forged.

Robin Rice

ORCID: <https://orcid.org/0000-0003-0214-6559>

University of Edinburgh, Library & University Collections

E-Mail: r.rice@ed.ac.uk

- 1 Finnie, Ellen and Arthur, Michael A. (2016). Being Earnest With Collections – Voting with our Dollars: Making a New Home for the Collections Budget in the MIT Libraries. *Against the Grain* 28(4), 52. <https://doi.org/10.7771/2380-176X.7496>
- 2 Cox, Amanda et al. (2017). Developments in Research Data Management in Academic Libraries: Towards an Understanding of Research Data Service Maturity. *Journal of the Association for Information, Science and Technology* 68(9), 2191. <https://doi.org/10.1002/asi.23781> [used with permission of publisher].
- 3 FOSTER, What is Open Science? Introduction: <https://www.fosteropenscience.eu/content/what-open-science-introduction>
- 4 Rice, Robin and John Southall (2016). *The Data Librarian's Handbook*. Facet Publishing: London, 152–153.
- 5 GO FAIR, FAIR Principles: <https://www.go-fair.org/fair-principles>
- 6 European Commission, HORIZON 2020 Online Manual – Funding & tender opportunities: http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm
- 7 Federer, Lisa (2018). Defining data librarianship: A survey of competencies, skills, and training. *Journal of the Medical Library Association* 106(3), 294–303. <https://doi.org/10.5195/jmla.2018.306>
- 8 Ayris, Paul et al. (2018). Open Science and its role in universities: a roadmap for cultural change (LERU Advice paper 24). <https://www.leru.org/publications/open-science-and-its-role-in-universities-a-roadmap-for-cultural-change>
- 9 Ayris, Paul et al. (2018). LIBER Open Science Roadmap. Zenodo. <http://doi.org/10.5281/zenodo.1303002>