

Review: *Research Data Management: Practical Strategies
for Information Professionals*

Martin Donnelly
DCC, University of Edinburgh

Abstract

A review of *Research data management: Practical strategies for information professionals*, edited by Joyce Ray, Purdue University Press, 30 Jan 2014, Paperback, 436 pp., ISBN-10: 1-55753-664-3, ISBN-13: 978-1-55753-664-8.

Received 30 May 2014 | Accepted 30 May 2014

Correspondence should be addressed to Martin Donnelly, DCC, Appleton Tower, Crichton Street, Edinburgh EH8 9LE.
Email: martin.donnelly@ed.ac.uk

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: <http://www.ijdc.net/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution 4.0 International Licence. For details please see <http://creativecommons.org/licenses/by/4.0/>



Beyond the rarefied strata of those for whom it is a central concern, research data management (RDM) can suffer from a bit of an image problem. Delicately put, it is a 'dry' topic on a good day, and an unwelcome, additional bureaucratic imposition on a bad one. Beyond that, it is very much a derived demand: just like taking out the trash or doing tax returns, few people actually want *to do it themselves*, but by and large they do tend to want it *to be done*.¹

Unlike taking out the trash or doing tax returns, however, data management – in universities at least – is a hybrid activity involving a multiplicity of business areas. Indeed, RDM cross-cuts so many functions and sectors – from academia to charities, government and industry – that it is a real challenge to speak to all parties in a single, engaging voice, and those of us who present frequently about RDM – to diverse and often recalcitrant audiences – hungrily accrue stimulating real-world exemplars in our quest to make the arcane and theoretical engaging and vital, and above all relevant to the often reluctant listener.

In the first chapter proper of this very welcome addition to the literature (Ray, 2014), Purdue Dean of Libraries James L. Mullins identifies and describes the multi-functional nature of modern computational science. There is a clear parallel here with the variety of stakeholder groups involved in RDM. If it takes a village to raise a child, it takes a university – or at least a broad cross-section thereof – to create, manage and preserve a dataset.²

Towards the end of 2013, a relative glut of RDM books emerged almost as one, and from diverse corners of the globe. From Australia, we got Simons and Richardson (2013) with *New Content in Digital Repositories*, while in the UK Pryor, Jones and Whyte (2013) gave us *Delivering Research Data Management Services*, and Corti, Van den Eynden, Bishop and Woollard (2013) *Managing and Sharing Research Data*. Now, with this book, Joyce Ray brings us the US perspective from a range of research libraries.

As the book's subtitle makes clear, this collection is directed towards an audience of librarians, archivists and other information professionals, working primarily in American universities; it is fun to ponder how well the strategies that are outlined here might travel. The data management space in US higher education is predominantly owned by the libraries (throughout the book there is an assumption that libraries are the 'natural' home for data management at universities), whereas here in the UK it is much more dependent on individual institutional cultures and circumstances whether it is the librarians, the academics, or the administrators who take the lead. It is not uncommon, for example, for a research funding office, or an IT service, to coordinate work in this area. (One wonders to what extent the US situation is related to the faculty status of academic librarians, and the aforementioned Mullins addresses issues of status and hierarchy in his case study here...)

So, what's in the box? After a very strong introductory chapter from Ray herself, we get seven themed sections. Part 1 (Understanding the Policy Context) offers chapters from Purdue's Mullins and UC Davis' MacKenzie Smith discussing the policy and institutional framework(s) that underpin RDM activities, and the technological environment within which they exist. Part 2 (Planning for Data Management) covers

¹ C.f. Shepley (2013).

² This is increasingly the view of research funders, notably the UK Engineering and Physical Sciences Research Council (EPSRC) which has introduced a policy identifying the institution, as opposed to the more short-lived research group, as the proper steward for data holdings.

lifecycle models, data management planning and assessment tools, repository audit and certification. Part 3 (Managing Project Data) covers copyright (including Open Data issues), metadata services and data citation. Part 4 (Archiving and Managing Research Data in Repositories) touches again on lifecycle models, partnership approaches and the possible uses of virtualized ('cloud') storage for cold storage of infrequently accessed data. Part 5 (Measuring Success) covers project evaluation and metrics, and Part 6 (Bringing It All Together) offers four institutional case studies (from Cornell, Purdue, Rice and Oregon). Finally, CNI's Clifford Lynch predicts and ranks the next group of challenges likely to be encountered in the curation of scholarly data.

Inevitably, as in any edited collection, some chapters are more satisfying than others (the case studies in particular are something of a mixed bag), and the ordering of content doesn't always flow as smoothly as it might. Occasionally we come across a statement late in the book which, if included earlier, would have changed your view of the overall message. Angus Whyte's paragraphs on aspects of data (p. 280), for example, would have been better placed earlier, as would the data citation chapter, citation being foundational in terms of 'the why'. (It is also a little disconcerting to find a paragraph headed 'What is research data?' 224 pages into a book dedicated to unpicking this exact subject.) More space is devoted to discussions of value for money and the need to meet funder mandates than, for example, gentler appeals to research quality and integrity; it is perhaps a cultural issue, but one feels that intellectual justification ought to take precedence over such worldly concerns as value for money.

Particular strengths? It is wholly admirable that as early as page 2 Ray is drawing attention to 'small science', and this focus sticks around throughout the entire book. The majority of National Science Foundation grants come in at \$1 million or less, averaging around \$200k, and as Jan Brase et al. point out:

the aggregated data produced by individual researchers or small research groups may well dwarf that created by [high-performance computing] or [scientific computing]. By its nature, these 'long tail' data are hard to find, standardize, and account for, but they still deserve proper data management.

Another topic addressed in multiple chapters here, and which seems likely to stand the test of time, is that of data management planning, and – wider than that – organising and tailoring your approach/infrastructure in order to streamline with existing working practices. This is increasingly considered as a pervasive, fundamental underpinning activity throughout the data lifecycle, and crossing from use case to use case. As Jake Carlson notes:

Appropriating a life cycle developed by another organization without careful consideration of its relation and fit to your local environment and target audience is a risky endeavor. The services you intend to provide should drive the life cycle model you create to structure and communicate these services to your intended audiences, rather than the other way around.

Here, too, is sage advice for data management planning: *caveat utilitor*. Andrew Sallans and Sherry Lake pick up this theme in the following chapter, where they describe narrow, requirements driven DMP support as 'teaching to the test' rather than inculcating firm understanding of the issues and shared benefits of virtuous data management habits.

There is also good, practical advice on legal aspects, which – while US academia-centric – should nonetheless be useful in other contexts.

With this volume, Ray has given us a good, practical primer for those already working in the information area (specifically librarians), written for an American academic audience but readily accessible for those beyond it. The range of issues covered is pretty comprehensive, although given the aforementioned hybrid nature of RDM it would have been worth having a dedicated chapter covering roles, training, advocacy, outreach, and coordination; that said, Mullins does touch on this briefly, and the Cornell case study contains helpful suggestions on how to constitute and organise institutional RDM teams. Training and education are also covered, briefly – and somewhat unintuitively – within the Metadata Services chapter). The text can become a little repetitive, with the same institutions and initiatives cropping up time and again, and some sections read like advertisements for particular technical solutions without much indication of how they compare against rival offerings. It also remains to be seen how the tools and techniques pioneered at advanced research universities and their libraries will translate to neighbouring sectors, or indeed vice versa.

At only £18 from Amazon (at time of writing), it offers great value for money, and Cliff Lynch's unflinching and to-the-point 'Motivations and Drivers' section should be required reading for doubtful colleagues, focusing as it does on pragmatic and operational imperatives for data management activities above more esoteric rationales.³ Not all of what is covered here is relevant in all contexts, but there is likely to be something for most audiences within it. If you're short on time, just reading Ray's introduction and Lynch's conclusion would provide any reader with a very decent snapshot of this developing discipline from the specialised viewpoint of the academic data librarian.⁴

References

- Corti, L., Van den Eynden, V., Bishop, L. & Woollard, M. (2013). *Managing and sharing research data: A guide to good practice*. London: SAGE.
- Pryor, G., Jones, S. & Whyte, A. (Eds.). (2013). *Delivering research data management services: Fundamentals of good practice*. London, UK: Facet.
- Ray, J. (Ed.). (2014). *Research data management: practical strategies for information professionals*. West Lafayette, IN: Purdue University Press.
- Shepley, J. (2013, November 20). Nobody really cares about information management. *CMSWire*. Retrieved from <http://www.cmswire.com/cms/information-management/nobody-really-cares-about-information-management-023234.php>

³ Lynch's closing chapter appears to be Open Access, and at the time of writing is freely available via the CNI website: <http://www.cni.org/wp-content/uploads/2013/10/Research-Data-Mgt-Ch19-Lynch-Oct-29-2013.pdf>

⁴ It remains to be seen, however, whether – and if so for how long – such specialised treatment is still necessary. 2013 was the year that data went mainstream, with omnipresent advertising references to 3G and 4G data plans, countless headlines about big data, and blanket media coverage of the covert – and at times seemingly indiscriminate – interception and storing of phone, SMS and email data/metadata by security/intelligence agencies.

Simons, N. & Richardson, J. (2013). *New content in digital repositories: The changing research landscape*. Oxford, UK: Chandos.