Book Review: The Data Librarian's Handbook by Robin Rice and John Southall

blogs.lse.ac.uk/lsereviewofbooks/2017/04/10/book-review-the-data-librarians-handbook-by-robin-rice-and-john-southall/

In **The Data Librarian's Handbook***,* **Robin Rice** *and* **John Southall** *examine the role of the data librarian, an emergent profession increasingly vital for academic libraries to support activities around Research Data Management (RDM). This is an accessible and engaging book full of interesting case studies and insights that will be essential for any information professional looking to broaden their knowledge of data management, writes* **Neil Stewart***.*

The Data Librarian's Handbook. **Robin Rice and John Southall. Facet Publishing. 2016.**

Find this book: amazon

This book is an overdue examination of that new, exotic addition to the librarian menagerie: the Data Librarian. Academic libraries and librarians have long had a role in managing access to data resources, and indeed LSE Library is this year celebrating the twentieth anniversary of its Data Library, a dedicated service for the provision of data for learning and research. However the increasing need for libraries to support activities around Research Data Management (RDM) has led to the creation of dedicated Data Librarian roles. It is this emergent profession that Robin Rice and John Southall's *The Data Librarian's Handbook* reflects upon and seeks to guide.

Now, you might be thinking that such a book is likely to be dry, technical and possibly slightly boring. I was worried about this too, but no such thoughts arose as I read it – it is well written, being both accessible for any general reader who might be interested and sufficiently sophisticated in its breadth of coverage and detail to provide insight for those professionally invested in the issues discussed. It helps that each chapter ends with a useful set of summary bullet points, as well as reflective questions for a couple of its target audiences: namely, students studying to become information professionals and early-career librarians who may be tasked with involvement in data-related services. There is also a set of interesting case studies from Data Librarian practitioners that provide local colour (full disclosure: one of these comes from LSE Library's very own Data Librarian, Laurence Horton).

One of the issues the book raises indirectly is the tension between Data Librarians as managers for the aforementioned data resources and Data Librarians as active participants in the research life-cycle when dealing with RDM matters. The former issue is addressed by some, in my view fairly commonplace, reflections on collection development that might have been left out with no great loss to the book as a whole. Much more interesting are the thoughts and advice on the practicalities of Data Librarians involving themselves in the research process, in particular using Data Management Plans (now often a requirement of research funders) as a way of introducing good RDM practice to researchers.
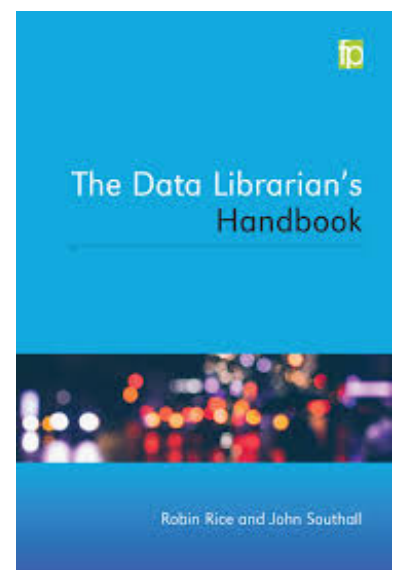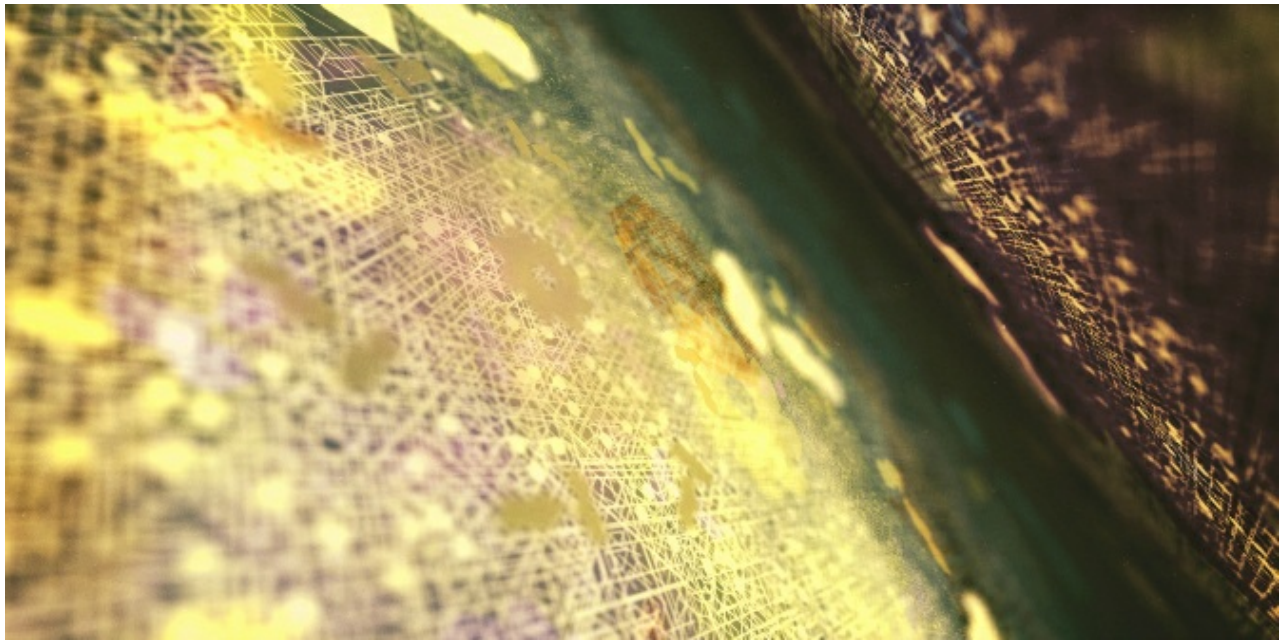
Image Credit: (beachmobjellies CC BY SA 2.0)

Another unacknowledged tension in the book arises over data repositories. Many UK universities have invested significant resources into creating institutional repositories for the preservation and serving of research data, and the authors provide good advice on how to go about setting up such a service. What is not examined is whether this is a good strategy for universities in the first place. LSE (admittedly a bit of an outlier as a specialist social science university) has not yet implemented such a repository, but it remains unclear whether this is a problem. If we build such a service, will researchers come? This problem is made more complicated by the existence of well-established domain-specific services such as the UK Data Service for the social sciences, and large, well-supported and visually arresting generic, open data repositories such as figshare and Zenodo.

On a more positive note, the book contains an excellent discussion of the issues surrounding the management of research data, particularly the problems that arise with sensitive or confidential data. There are many circumstances in which research data may fall into these categories, but the authors also identify what might be called the 'special flower' problem: researchers, for very good reason, often get attached to what they consider to be 'their' data (in contrast to research papers that arise from a research project which are often considered tombstones for completed projects). This happens for a number of legitimate reasons: for example, there may be a wish to glean further insights from the data, or a belief that the dataset won't be comprehensible outside of their research group without onerous and boring documentation. As a result, it can become difficult to persuade researchers to archive their data in a repository, let alone consider sharing it. These barriers to archiving are an aspect of researcher behaviour that have been more thoroughly examined elsewhere.

Looking to the future, Rice and Southall examine the ways in which RDM and data librarianship contribute to wider discussions about open scholarship and open science. They examine data reproducibility as a method of validating the scientific record, and note that certain disciplines such as the social sciences have a long way to travel in this regard. They recount the depressing stories of scientific fraud in the Netherlands earlier this decade, which shone a harsh light on the practices of the social psychologist Diederek Stapel: the so-called Stapel Affair. More encouragingly, they note recent moves to treat research datasets as first-class research objects, worthy of citation and (crucially) academic credit. The requirements for the forthcoming REF 2020 here in the UK are likely to help spur this change, with archiving and making research data available finally due to receive the respect they deserve.

The Data Librarian's Handbook covers a large amount of interesting terrain in thoughtful and accessible ways. It is both essential for any information professional interested in data and their management, and is also indicative of the

increasing – and increasingly varied – role that data and data management play in libraries and more broadly across academia.

Neil Stewart is the Digital Library Manager at LSE Library. He manages LSE's Digital Library, an online repository of digitised and born digital materials from LSE Library's rich collection of social science holdings. He is interested in digital scholarship, digitisation for scholarly re-use, research data management, open access, open science and web technologies for libraries. Neil holds Master's degrees in International Relations (Manchester) and Library and Information Science (UCL).

Note: This review gives the views of the author, and not the position of the LSE Review of Books blog, or of the London School of Economics.