

Academic librarianship as a data profession: The familiar and unfamiliar in the data role spectrum

Andrew Cox, Senior Lecturer, Information School, University of Sheffield

a.m.cox@sheffield.ac.uk

Summary

The paper reflects on the growing importance of data in the work of academic librarians. Based on tasks related to research data management (RDM) it proposes a data role spectrum, ranging from familiar tasks such as supporting search, training and collection management, through to less familiar tasks, such as data analysis and visualisation. Areas where library work is increasingly tied up with data, such as Text and Data Mining and bibliometrics, can be analysed for their positions on the spectrum.

We are used to thinking of data as sitting beneath information and knowledge in a pyramid of value. But the concept of “data” has grown significantly in importance in the last five years, driven by public speculation about the power and risks of big data. In parallel, data related roles are becoming increasingly important in professional practice. Librarianship is becoming more of a “data profession”. One obvious starting point for thinking about this impact in the academic library context is to analyse the specific case of Research Data Management (RDM). This is an area where dealing with data is clearly central. But is this familiar territory, just reinventing or extending what librarians already do or does it require a new set of competencies?

The most direct way to understand what Research Data Management means to academic librarians in practice is to think of some of the key tasks that are involved in delivering a research data service. If one were to make a shortlist it would certainly include: helping a researcher to find pre-existing data sources relevant to their research; running a training or awareness session; reviewing metadata associated with a potential deposit into a data repository; investigating what researchers need in terms of support; inputting to the creation of a data policy framework; and offering advice on a Data Management Plan (DMP) for a project proposal. Nearly all of these activities have strong continuities with what we already expect to do as a librarian.

What could be more natural for a librarian to find themselves helping someone search for and evaluate a source, albeit data rather than a published text? It is true there are a few differences from supporting the usual literature search tasks. One needs to know a bit about what data sources and data archives there are out there. There might well be licence conditions associated with data reuse that need careful analysis, in a way not applicable to published texts. Further, a source of data might not necessarily be ready packaged: it could be that there is an API or some data service through which the

researcher can generate a relevant dataset. However, overall this feels like a very familiar role that a librarian would have the requisite skills for (Gregory et al., 2018).

Adding something about RDM to an existing user training session is again the familiar ground of academic librarianship, and its information literacy focus. In all likelihood, the objectives of such a session would be to make researchers aware of institutional, funder and publisher requirements; to make the case for data sharing, albeit qualified by a sense of potential exceptions; and explore practical data management basics (such as back-ups and file-naming). So it would require acquiring new knowledge, yet there are strong continuities with open access advocacy, and learning outcomes around these topics would fit into existing information literacy sessions for researchers.

Reviewing a potential deposit to the data repository is again familiar territory. It is a fairly standard library task rooted in collection management principles and based on an understanding of the importance of metadata and standards. It might well be combined with the role of monitoring metadata related to outputs in the repository.

Gathering requirements from researchers about their support needs also should come naturally to a user service focussed profession like librarianship. Librarians are used to gathering data from interviews, focus groups and surveys to try and discover what services users need, and then designing services or procuring systems to meet these needs. RDM takes us deeper into the research process as an aspect of scholarly communication, but a strong interest in user behaviour is a good starting point for carrying through this task. Data policy is about creating a governance structure within which data is valued and managed. Contributing to the development of such a policy is probably a fairly familiar task, and requiring a good understanding of the wider policy context, such as relevant institutional and national policies.

Helping a researcher write their DMP could be the most unfamiliar of all the tasks I listed above. While the skills needed to do it effectively are the same as any advice service, this particular support requires a fairly deep knowledge of funder requirements; of relevant standards; of local data management processes (e.g. around data storage); as well as a feel for the research process.

Thus so much of what RDM is about could be considered somewhat familiar territory. It seems to involve acquiring new knowledge but much of the role is familiar. There are plenty of ways that roles in RDM build on skills and knowledge that most librarians already have (Cox and Verbaan, 2018). However, there are some more tasks that could potentially be involved in supporting RDM. If we look at the other end of the spectrum from the common tasks, we could also identify potential roles in data curation, data carpentry, data integrity, data analysis & visualisation and also embedded roles in research project teams. These are more like the specialist or cutting edge of RDS. They would include:

- Data curation is the long-term digital preservation of datasets. Preservation has traditionally been an aspect of library work, but it is perhaps more in the territory of archives.
- Data carpentry is about understanding how to manipulate and transform data, preparatory to analysis.
- Data integrity links to the traditional interest of information professionals in data quality, but in the context of a crisis of reproducibility (in certain subjects), the assurance perhaps lies with more open science.
- Embedded roles involve working directly with a research team. This is more about how the role is delivered than the knowledge/expertise required. It is about breaking out of the library and working with researchers on a daily basis.
- Supporting or undertaking data analysis and visualisation, or at least having a role in the selection and supporting use of computational tools to do analysis.

It remains to be seen whether academic libraries will start to see these roles as standard tasks. Probably institutions will vary depending on their research-intensive nature, among other factors. Emerging from this discussion is what I am calling the “data role spectrum”. The data role spectrum starts with the familiar and transitions to the unfamiliar.

- Support for data search/access to data
- Data literacy training and promoting awareness
- Data collection management, including metadata
- Gathering support requirements for services/tools
- Data policy
- Data management advice
- Data carpentry
- Data curation
- Data integrity
- Embedded roles in a research team
- Data analysis and visualisation

At one end of the scale are the data related tasks that feel relatively familiar, such as data search, data literacy training and data collection management. At the other end of the scale are those that we might associate with the work of IT professionals, such as data carpentry, or researchers, such as data analysis and visualisation. These fit less easily into the classic library role, but the profession is changing to engage with them more. I think there could be very different perceptions of how “familiar” some of these roles are and there is a lot to unpack under the heading of something like data curation, but the spectrum is useful for looking at other areas where the “deluge of data” is flooding into academic library work. Here are a few examples. One emerging area is Text and Data Mining (TDM) where machine-learning tools are used to analyse huge unstructured corpuses of texts. In TDM the library role is most likely to be about licensing content (data collection management) and training people to use tools (data training), and in managing derived outputs (a form of data curation). Another area that could be analysed through the spectrum is around bibliometrics and altmetrics, where “the data” is about researchers and their outputs, analysed to reveal the impact of research. Here the role seems mostly to be about calculating metrics, and helping researchers to understand

metrics about themselves and their work, in the context of the concept of responsible use (Cox et al. 2017; Bibliomagician, 2017). Interestingly, here the role is actually in the area of doing analysis/visualisation, albeit, through proprietary tools, and with an emphasis on training others to undertake analyses for themselves. One could do a similar review for other areas such as library and learning analytics, although here it is about analysing data to support management decisions, rather than supporting users.

Academic librarianship seems to be moving towards becoming a data profession. It will be interesting to see how things develop over the next decade. Perhaps data analysis will become integral to professional competencies. This would be a fundamental shift in the positioning of academic librarianship. There are much more obvious areas where including data in support services or training is a natural extension of what academic libraries already do; and other areas such as data curation that others would expect librarians to take up. The data role spectrum could help chart these changes.

References

[Bibliomagician](#) (2017). *Competency model for bibliometric work*.

Cox, A.M. and Verbaan, E. (2018). *Exploring Research Data Management*. London: Facet.

Cox, A., Gadd, E., Petersohn, S., & Saffi, L. (2017). [Competencies for bibliometrics](#). *Journal of Librarianship and Information Science*.

Dooley, J. M. (2015). *The archival advantage: Integrating archival expertise into management of born-digital library materials*. Dublin, OH: OCLC Research.

Gregory K, Khalsa SJ, Michener WK, Psomopoulos FE, de Waard A, Wu M (2018) [Eleven quick tips for finding research data](#). *PLoS Computational Biology* 14(4)

Andrew Cox and Eddy Verbaan's UKeIG CPD course, [Research Data Management for Information Professionals](#), is on Wednesday, 14th November 2018 at CILIP headquarters in London