IJDC | *Peer-Reviewed Paper*

Progress in Research Data Services: An International Survey of University Libraries

Andrew M. Cox
University of Sheffield

Mary Anne Kennan
Charles Sturt University

Elizabeth Josephine Lyon
University of Pittsburgh

Stephen Pinfield
University of Sheffield

Laura Sbaffi
University of Sheffield

Abstract

University libraries have played an important role in constructing an infrastructure of support for Research Data Management at an institutional level. This paper presents a comparative analysis of two international surveys of libraries about their involvement in Research Data Services conducted in 2014 and 2018. The aim was to explore how services had developed over this time period, and to explore the drivers and barriers to change. In particular, there was an interest in how far the FAIR data principles had been adopted.

Services in nearly every area were more developed in 2018 than before, but technical services remained less developed than advisory. Progress on institutional policy was also evident. However, priorities did not seem to have shifted significantly. Open ended answers suggested that funder policy, rather than researcher demand, remained the main driver of service development and that resources and skills gaps remained issues. While widely understood as an important reference point and standard, because of their relatively recent publication date, FAIR principles had not been widely adopted explicitly in policy.

Received 06 December 2018 ~ *Revision received* 13 August 2019 ~ *Accepted* 13 August 2019

Correspondence should be addressed to Dr Andrew Cox, Rm 222, Information School, Regent Court, Portobello, The University of Sheffield, Sheffield S1 4DP. Email: a.m.cox@sheffield.ac.uk

An earlier version of this paper was presented at the 13th International Digital Curation Conference.

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/

Introduction

Increasing involvement in research data management (RDM) has been one of the major changes in academic library work in the last decade. Intense interest has developed in understanding researchers' data practices, identifying their support requirements and building policy and services to meet their needs. Libraries have been heavily involved in developing RDM policies at the individual institutional level. Many have also recognised a need to provide a range of training and advisory services, e.g. for data management planning. In addition, technical services often built around a research data repository and data preservation, and sometimes around data analysis and visualisation, have emerged. Delivering these services has demanded recruitment of staff, reskilling of existing staff and some organisational restructuring. In this context, there is a strong interest in benchmarking Research Data Services (RDS) and understanding change in the sector as a whole.

Seeking to characterise the emergence of RDS, a number of surveys have been conducted at national and international level (Tenopir, Birch and Allard, 2012; Corrall, Kennan and Afzal, 2013; Cox and Pinfield, 2014; Whyte, 2014; Cox, Kennan, Lyon and Pinfield, 2017; Tenopir et al., 2017). Several common themes emerge from such studies. Typically, it is what Tenopir et al. (2017) dub 'consultative services' that are more common than either technical or hands-on services. It is usual for RDM support to be a collaboration between different parts of the institution, especially the library, IT and research administration. Bryant et al. (2017) argue strongly that different institutions will offer different packages of service based, among other factors, on their local needs and resources.

One of the most recent surveys conducted by some of the current authors in 2014, revealed a complex international landscape of academic library responses (Cox, Kennan, Lyon and Pinfield, 2017). There were indications of significant leadership activity from the library community in, for example, coordinating and promoting the development of policies and initiating research data audits or assessments. However, there was less evidence of mature services such as technical support and infrastructure, data curation, or project participation through embedded or immersive roles. A tentative maturity model was produced from the data to capture the range of current service developments internationally.

Qualitative studies give us a sense of factors shaping librarians' involvement in RDM. For example, Faniel and Connaway (2018) identify a number of key areas both facilitating and constraining their involvement, including: technical resources, human resources, researchers' perceptions, leadership support and collaboration. Technical complexities of managing research data, lack of sufficient expertise and researcher assumptions about the library were all seen as major barriers to RDS. Many of these shaping factors are also identified in an earlier study, which also identifies jurisdictional issues around the role of the library and other institutional services, such as IT (Verbaan and Cox, 2014; Pinfield, Cox and Smith, 2014).

In thinking about what type of services are needed, the FAIR data principles (findable, accessible, interoperable and reusable) offer a concisely articulated set of ideals (Force 11, 2016) around which services and resources could be based. As a new initiative in the sector, FAIR is quickly gaining traction, though it may be open to different interpretations and awareness among researchers seems to still be patchy (Allen and

 doi:10.2218/ijdc.v14i1.595

Harland, 2018). It was of interest to try to discover how this and other principles and standards are influencing current support practice.

The aim of this paper is to update this analysis of the character of RDS development in libraries internationally; more specifically it seeks to answer the following four research questions:

1. How, if at all, have the types and levels of RDS being offered changed?
2. Has there been any change in the types of service that are perceived to be a future priority?
3. What have been the main drivers and barriers to developing RDS and policy?
4. What has been the influence of the FAIR principles to date?

The findings are based on a survey of libraries conducted in 2018, and analysed comparing the results to the 2014 study (Cox, Kennan, Lyon and Pinfield, 2017).

Methodology

Given the aim to capture a picture of the international development of RDS, and to analyse change over time, the appropriate methodology was to essentially repeat the web-based survey of our previous study. The revised survey was similar in seeking to explore libraries' perspectives on the development of policy, requirements gathering activity and collaboration, but particularly to ask about current services offered and future service priorities. A few changes were made partly to reduce the overall number of questions but also to ensure the survey's currency, e.g. to ask about principles or standards governing policy and practice, such as FAIR. We also added an open text question about drivers to complement an existing one on challenges. A redesigned survey consisting of 24 questions, hosted on SurveyMonkey, was piloted, and then distributed between February and April 2018.

As in the 2014 survey, the countries surveyed were Australia, Canada, Germany, Ireland, Netherlands, New Zealand and the UK. All the universities in these countries were invited to participate. In addition, a request to participate was extended to a smaller set of institutions in the USA: the Association of American Research Libraries (ARL). We sought one reply per institution by inviting library directors directly by email, except in the case of Germany and the USA, where a more broadcast approach was taken. 209 responses were received in total. Table 1 shows that the response rate from Australia, Ireland and New Zealand was high. Around 50% of UK institutions responded, comparable to 2014. There was a good range of types of institution in the UK response. Response rates from other countries was lower and this combined with the small numbers of institutions in some countries reduced the potential for statistical analysis. There is also likely to have been a non-response bias, with institutions with a strong investment in RDM more likely to reply, and so a more positive picture of sector progress than actually the case is likely to appear in the survey. Nevertheless, we believe that the study gives a good guide to current trends in RDS.

Table 1. Response rate.

Country	Responses	Number invited	Response rate	2014 response
Australia	34	39	87%	34
Canada	24	74	32%	17
Germany	23	c250	c10%	6
Ireland	11	12	92%	7
Netherlands	6	16	38%	12
New Zealand	8	8	100%	7
UK	80	169	47%	85
USA	23	86	27%	n/a
Total	209			

The data from the questionnaire was analysed through descriptive statistics. In addition to closed questions, a substantial amount of data was collected in the form of text responses to open-ended questions, principally about drivers and challenges to development, amounting to 15,000 words of data. This text was analysed through content analysis. Selected findings are presented here.

Findings

As Figure 1 reveals, the pattern of policy development was quite varied.

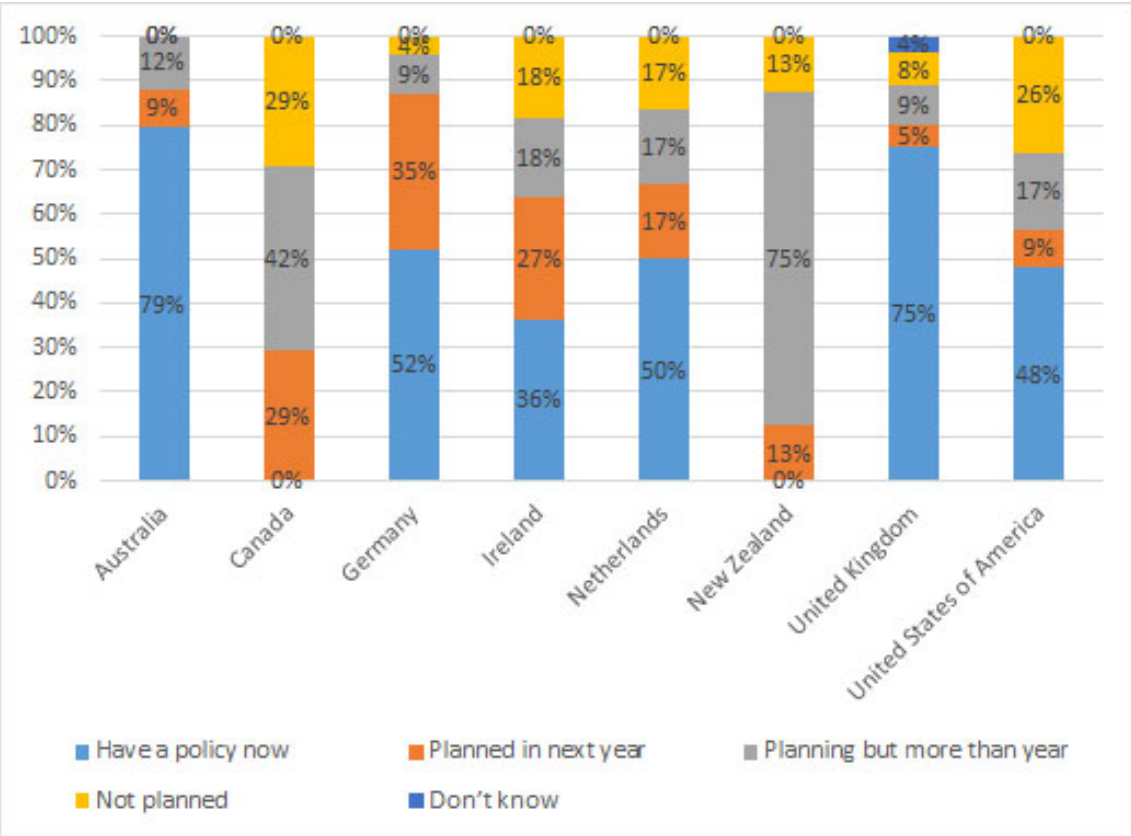

Figure 1. Formal RDM policy in place or planned.

117 (56%) of responding libraries stated that they had a formal RDM policy. In addition, 29 (14%) were planning to have one in the next year and a further 36 (17%) in the more distant future. Only 11% had no plan to develop a policy. Yet there is a diverse picture internationally. Three quarters of Australian and UK institutions did have a policy already, whereas none of the libraries who participating in the survey in Canada or New Zealand did. Comments in the survey suggest that significant changes are on the horizon for Canada and Ireland, like a number of other European countries (SPARC/DCC, 2018). In 2014, 64 out of 167 institutions (38% of those surveyed) had had a policy in place, with 76 (46%) having plans for one. This indicates "progress" across the sector, but it does suggest that not all the institutions who said they were planning a policy in 2014 did implement one, assuming both sets of responses are representative.

Probing about involvement in policy and service development showed that librarians, research administrators often played a leadership role in policy development; however, it was more typically libraries that led on service development. It was also usual for IT services and academics to be involved in policy and RDS development, but they did not seem to adopt a leadership role.

Based on treating responses self-reported ratings of "no service" = 0, "basic service" = 1 and "well developed or extensive service" = 2, Figure 2 offers a clear visualisation of the progress made in developing RDS, but without a major shift in emphasis away from advisory services. Note some items were not available as options in 2014.

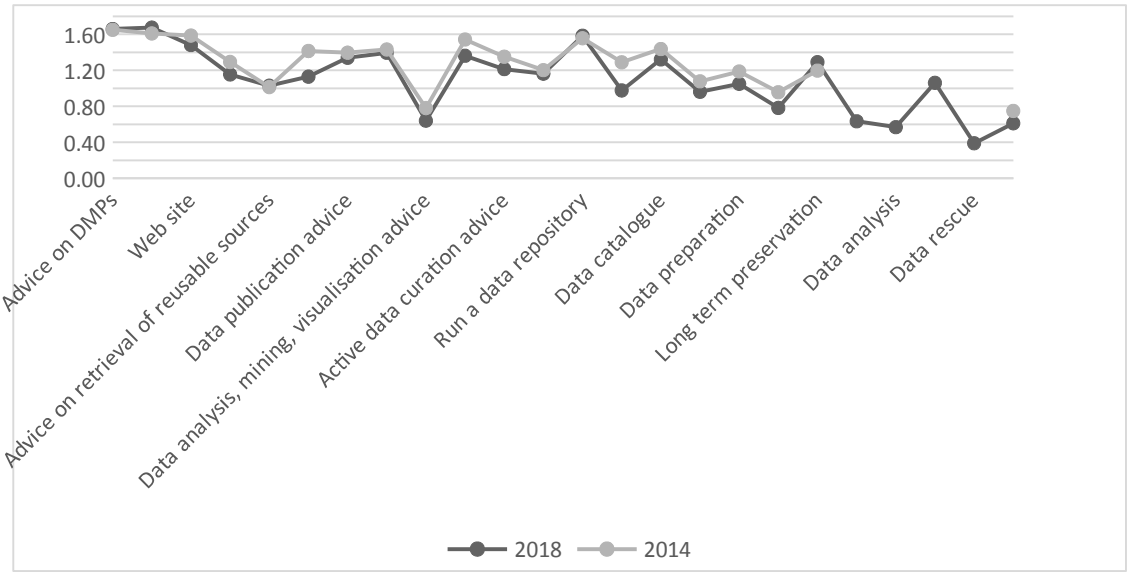

Figure 2. A comparison of RDS between 2014 and 2018.

As in 2014, advisory services are the main types of service offered by libraries. These include help with data management planning (DMP), web guides, data discovery and support on copyright. Nine of the top ten library based RDS were "advisory" (as categorised in the survey). Running a repository was the only "technical" service in the top ten. Advisory services on data analysis, data mining and services for cleaning data, analysing and visualising data and rescuing legacy data remained relatively rare. In a small number of cases, the wording of questions had been changed, reducing the opportunity to compare. The 2014 question about directly participating in research projects had been substantially reworded in 2018 to mention 'embedding'. This resulted

in a fall in the numbers of people saying they had such a service, although it is possible that this was simply the result of interpretation of the phrase "embedding".

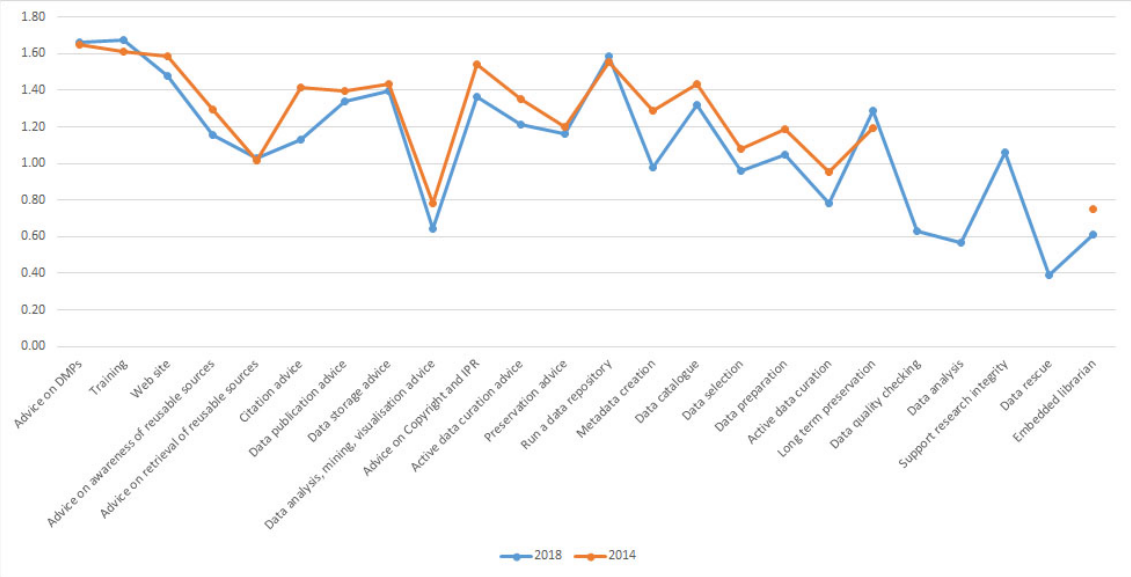

Figure 3. A comparison of RDS priorities 2014 and 2018.

Figure 3 suggests relatively little change in priorities between the two studies. In a couple of areas, such as data citation advice and metadata creation, services seemed to be slightly less of a priority than they were in 2014. However, overall the pattern is strikingly similar, strongly suggesting no significant change in aspirations around RDS.

Our understanding of the underlying drivers in the situation can be informed by the qualitative data from the survey. 167 of respondents wrote something about drivers in the open text box on the survey. Of these 57% identified compliance with funder policy as a driver for RDS. Libraries' perception of a potential role for themselves was another frequently mentioned driver. Sometimes this implied that libraries had the right skill set to contribute, in other accounts it seemed more that the library was searching for a role.

> "It is a natural extension of the 'traditional library tasks'."

> "The need to expand the services we offer to keep the library 'current'."

Researcher need was recognised as important, but this was often couched in terms that suggested that researchers did not recognise their own needs. FAIR was only explicitly mentioned at this point in the survey in six responses. However, when asked directly in another question about guiding principles for service delivery (which gave FAIR as an example), there was a strong sense that it was becoming a well-recognised set of principles among librarians. Many answers expressed that respondents thought their service did align to FAIR. However, not surprisingly given that FAIR is relatively new, explicit mentions in policy were relatively rare. Only where the policy was being refreshed or where a new policy being created was it going to be explicitly tied to FAIR.

> "We are currently reviewing our policy with the aim of integrating FAIR principles into the policy."

 doi:10.2218/ijdc.v14i1.595

> "No principles are explicitly referenced in the policy but FAIR has informed our approach. We do reference these principles in our training and advocacy activities."

From this perspective the fact that so many libraries were still planning a policy was positive for the widespread acknowledgement of FAIR. It was apparent that explicitly referenced principles in policy were much more likely to be national level standards, such as in Australia's "Code for the Responsible Conduct of Research" or in the UK's "RCUK Common principles" and "Concordat on Open Data".

170 respondents wrote something about challenges. Here, lack of researcher engagement was cited as one major barrier. Thus, it seems that RDM remains premised on a top-down mandate, rather than strong demand from researchers themselves. The most frequently mentioned challenges, however, were skills and resources:

> "Staff skills and willingness to take on new tasks that are not viewed as traditional 'library' tasks."

> "Resources and time, there are many areas we are being pulled into but we do not have the staffing or relevant expertise."

> "A major challenge is doing this as well as everything else. Also, RDM is much more complex than most other things we do."

Such challenges interact with each other to create a sense that further development is blocked. For example:

> "The chicken and egg scenario of RDM remains. You need to have a service in place to promote effective RDM practices, but it is hard to fund and develop a service without evidence of demand for that service, or to decide how to scope it. We are still in advance of academic demand for RDM."

Data from the study indicated a marked, if declining skills gap. Data curation skills, knowledge of research methods, data description were the most commonly cited areas where more skills were most needed in the library. But nearly all the options offered, including technical and ICT skills, knowledge of the research lifecycle, legal and policy knowledge and understanding of research integrity, were perceived to be knowledge and skills still needed by over 50% of respondents.

To reinforce the sense that the practice of RDM support is still in development, multiple models of organisational structure seemed to have roughly equal popularity: while some vested the role in a single individual (23%), others did so through a general research support team (19%), others in a dedicated RDM team (23%).

Discussion

The data provides evidence of broad "progress" internationally in the creation of RDS by libraries. In nearly every area of activity more institutions are delivering a service, and these are more often at the well-developed or extensive level. This pattern applies to both advisory and technical services, though technical services remain less developed. 67% of institutions have a data repository or store; many advisory services are provided at some level at above 75% or even 80% of responding institutions. The extent to which the number of institutions have been developing policies seems less marked, with a large number of institutions remaining in the planning stage for this.

It is interesting, however, that priorities were largely the same in 2018 as they had been in 2014. This suggests that the underlying agenda is basically unchanged. It may be that the growing momentum around open scholarship will lead to a significant re-evaluation of priorities, but there was no evidence of this yet in the survey results.

While one could think of the maturing of RDS internationally, there are significant gaps remaining. Less than a quarter of libraries are involved in offering services around data analysis and data mining. Such services are sometimes supplied by other parts of the institution, it is a minority of institutions that appear to offer a service. This may reflect that it lies too far outside libraries existing roles, expertise and perceived jurisdiction (Verbaan and Cox, 2014; Pinfield, Cox and Smith, 2014) to be easily accommodated. Further, a significant proportion of institutions do not have a policy in place; indeed, around 10% have no plan for a policy.

Qualitative data from the survey strongly indicates that the main driver for creating RDS remains compliance with funder policy, rather than researcher demand. Lack of researcher engagement was cited as a major barrier; combined with lack of resources and skills. These findings resonate strongly with Faniel and Connaway's (2018) characterisation of the factors shaping librarians' ability to respond to the RDM agenda. They too found researcher attitude, skills and resources as constraints in service development.

The data suggest that the dominant factor continuing to constrain RDS development is the strength of policy commitment. In the UK, for example, it is not clear how far compliance will be actually enforced. Canada and New Zealand seem to be on the verge of another phase of development, stimulated by national policy change. We should not assume that further development is necessarily desirable, however, it seems reasonable to argue that without an incontrovertible mandate it seems that RDS development will only progress gradually. The activity around FAIR arguably represents a more grassroots culture change, but although FAIR was recognised as relevant it was only just beginning to gain ground in explicit policy. The survey results suggest that libraries are still some way from addressing the more technical services beyond FAIR, which have been characterised as the FAIReR and FAIReST concepts proposed by Lyon (2018).

In reflecting on the findings it is important to acknowledge the potential for there to be more than one model of maturity. Less research intensive institutions may have less strong needs for a deep level of service, and the differential pattern across the data suggests that national level patterns are quite different from each other.

Conclusion

Libraries are often leading development of support infrastructure for RDM, which has been one of the most dynamic areas of academic library development in the last decade. Studies of this pattern of development are both significant for understanding how research is changing, and more specifically for understanding the changing role of academic libraries. This survey offers a significant insight into both subjects; and is the first to give a clear picture of change over time. Through rigorous analysis by descriptive statistics and content analysis of qualitative data the study provides a clear picture of how RDS are developing internationally. Nevertheless, a more complete picture would emerge from surveying research administration and IT departments who have also been strongly involved in RDS and might view developments somewhat differently. A more comprehensive response from Germany, USA and Canada might also change the impression of how RDS are developing. It would also be interesting to cross compare developments and progress in the different countries with the existence of national RDM policies explore developments in other countries, for example in other parts of Europe and in the global South, to see how far the same sort of patterns are emerging.

An anonymised version of the data on which this study was based has been shared via the University of Sheffield data repository, ORDA (Cox et al., 2019).

References

Allen, R. & Harland, D. (2018). Fair in practice: Jisc report on the Findable Accessible Interoperable and Reusable Data Principles. doi:10.5281/zenodo.1245568

Bryant, R., Lavoie, B. & Malpas, C. (2017). The realities of research data management: Part two: Scoping the university RDM service bundle. Dublin (OH): OCLC. doi:10.25333/C3Z039

Corrall, S., Kennan, M., & Afzal, W. (2013). Bibliometrics and research data management services: Emerging trends in library support for research. *Library Trends, 61*(3), 636–674.

Cox, A.M., & Pinfield, S. (2014). Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science, 46*(4), 299–316. doi:10.1177/0961000613492542

Cox, A.M., Kennan, M.A., Lyon, L., & Pinfield, S. (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*. doi:10.1002/asi.23781

Cox, A.M., Kennan, M.A., Lyon, L., & Pinfield, S. (2019). International survey of research data management in libraries. Retrieved from https://figshare.shef.ac.uk/articles/International_survey_of_research_data_management_in_libraries/9204509/1

Faniel, I.M. & Connaway, L.S. (2018). Librarians' perspectives on the factors influencing research data management programs. *College and Research Libraries, 79* (1) 100-119.

Force 11. (2016). The FAIR data principles. Retrieved from https://www.force11.org/group/fairgroup/fairprinciples

Lyon, E.J. (2018). Stewardship and Science: Reflections on Data-driven Opportunities for Libraries. Gottingen-CODATA RDM Symposium, Gottingen, March 2018. Retrieved from https://drive.google.com/file/d/1CNHXCO3q1R8WpeHt-jvr71o1IBwgWHKB/view

Pinfield S., Cox A.M., & Smith, J. (2014). Research data management and libraries: Relationships, activities, drivers and influences. *PLOS ONE* 9(12): e114734. doi:10.1371/journal.pone.0114734

SPARC/DCC. (2018). An analysis of open data and open science policies in Europe, v2.1. Retrieved from https://docs.google.com/document/d/1bC7EHsq6yplVKti6HMgKVhaR3T0qfRMwe2oSsej1xs0/edit

Tenopir, C., Birch, B., & Allard, S. (2012). Academic libraries and research data services: Current practices and plans for the future: An ACRL white paper. Chicago, IL: Association of College and Research Libraries. Retrieved from http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf

Tenopir, C., Talja, S., Horstmann, W., Late,E., Hughes, D., Pollock, D., Schmidt, B., Baird, L., Sandusky, R. & Allard, S. (2017). Research data services in European academic research libraries. *Liber Quarterly, 27*(1).

Verbaan, E., & Cox, A.M. (2014). Occupational sub-cultures, jurisdictional struggle and third space: Theorising professional service responses to research data management. *The Journal of Academic Librarianship, 40*(3-4), 211-219.

Whyte, A. (2014). Final results from DCC RDM 2014 survey. Retrieved from http://www.dcc.ac.uk/blog/rdm-2014-survey