Review Article

Indian Journal of Library and Information Science
Volume 13 Number 2, May - August 2019
DOI: http://dx.doi.org/10.21088/ijlis.0973.9548.13219.8

Research Data Management and the Role of Library Professionals

M. Natarajan[1], Manju Dahiya[2]

How to cite this article:

M. Natarajan, Manju Dahiya. Research Data Management and the Role of Library Professionals. Indian j. lib. inf. sci. 2019;13(2): 119–126.

[1]Consultant at Indira Gandhi National Centre for the Arts (IGNCA), Kala Nidhi Division, New Delhi 110011, India & Ex-Senior Principal Scientist, CSIR-NISCAIR, New Delhi 110001, India. [2]Semi Professional Assistant, Mata Sundri College for Women, Mata Sundri Lane, New Delhi 110002, India.

Address for correspondence

M. Natarajan, Consultant at Indira Gandhi National Centre for the Arts (IGNCA), Kala Nidhi Division, New Delhi 110011, India & Ex-Senior Principal Scientist, CSIR-NISCAIR, New Delhi 110001, India.

E-mail: drnatarajanm@gmail.com

Received on 12.06.2019,

Accepted on 11.07.2019

Abstract

Data Management is becoming increasingly important to researchers in all fields. The different types of data are explained with the need for management and planning of data. It explores the process of research data management and the services. The research data services and the components like research data planning, meta data service, storage, sharing, and reference service and data literacy are discussed. Libraries have been providing services to researchers, academicians and students for many years. Changes in research and the creation of data have given opportunities for libraries to be more involved in the context of research data management. The capability of the librarian, the need for staff development and skills that are needed to manage research data are discussed.

Keywords: Research data management; Data management; Data management planning; Sharing; Metadata; Research data services.

Introduction

The digital transformation in scientific research towards a data-intensive, collaborative approach has been well documented and discussed everywhere. There are challenges for the research community due to the advancement of digital technologies that has strengthened the power and reach of data. Organizations also face many other challenges while attempting to preserve the vast amount of data for long-term use. It is to describe data in a consistent way and also to share effectively the data with evolving data standards consistently. NRC (2009)[7] emphasized the sociological obstacles to data sharing and data reuse, with the huge increase in the amount of data being created as it was already emphasized over the past decades. Tenopir *et al.* (2012)[13] argued that, as science grows and moves towards a more collaborative, data-intensive and computational research, researchers are faced with various data management needs. The challenges surrounding research data has been assisted by the academic libraries. Tony Hey and others, (2009)[14] called this new data-intensive research environment of scientific study as the "fourth paradigm" of scientific inquiry. Many research funding bodies including the National Science Foundation (2008) recognize the importance of providing services and infrastructure to organize and preserve research data. Academic research libraries are located to base these research data services (RDS). Association of Research Libraries (2010) identified that the academic library is currently working to develop RDS as a new set of strategic services. To have RDS,

researchers are supposed to have the ability to collect, analyze, share and effectively manage and preserve research data. In many cases it has been found to be lacking in their data management, both short and long-term. Tenopir *et al.* (2011)[13] found that lack of institutional guidance and support, insufficient or nonexistent training and tools, and inadequate funding are the major barriers to data sharing by scientists, which may cause a loss or misuse of research data. Librarians can better satisfy the needs of researchers and are powerful in the new data-intensive research atmosphere. At this early stage, one has to identify those librarians who are working in academic research libraries and are actually engaged in providing RDS in particular research data services. This article provides an idea of research data management and the role played by the library professionals.

Data and categories

Data has been defined as the recorded information, regardless of form or media on which it may be recorded. It includes computer software and records of scientific and technical nature. Scientific data include both intangible data (statistics, findings, conclusion, etc) and tangible data that are not limited to notebooks, printouts, electronic storage, photographs, slides, negatives, films, scans, images, auto radiograms, etc. The investigator has the clear responsibility for recording, retaining and storing research data (Univ. of Pittsburg, RDM Guidelines).

NAP (1999) defines data as "facts, numbers, letters, and symbols that describe an object, idea, condition, situation, or other factors". The notion of "data" can vary considerably among collaborators and even more so between disciplines. Features of data have both immediate and enduring value, some gain value over time, some have transient value, and yet others are easier to recreate than to curate (Wallis JC and others, 2008)[15].

Johns Hopkins University, Baltimore, USA defined the research data as "a record of the results to be used to reconstruct and evaluate reports or otherwise published, such as laboratory notes, original experimental results, and instrument output values". Yale University, Connecticut, USA defines research data as "information collected, observed, or created for analysis purposes to produce the original study". Varieties of research data can be of observation variables like survey data on rainfall, wind speed, and water quality, seismic simulation data, laboratory data, and derived and compiled data for text mining or testing algorithms. The file formats for the research data can be of any digital file such as video, text, photos, numbers, etc.

The data has been categorized into three by National Science Foundation viz. i) observational data ii) computational data and iii) experimental data.

- Observational data refers to the data that are observed / include weather measurements and altitude surveys, either of which may be associated with specific places and times or may involve multiple places and times (e.g., cross-sectional, longitudinal studies).
- Computational data resulting from executing a computer model or simulation, whether for physics or cultural virtual reality.
- Experimental data from laboratory studies such as measurements of chemical reactions or from field experiments such as controlled behavioral studies. Whether sufficient data and documentation to reproduce the experiment are kept, varies by the cost and reproducibility of the experiment (NAP, 1995).

Data Management

Data Management includes acquiring, validating, storing, protecting, and processing required data to ensure the accessibility, reliability and timeliness of the data for its users. Big Data is more used by organizations and enterprises to inform business decisions and gain deep insights into customer behavior, trends, and opportunities for creating extraordinary customer experiences. Data Management (DM) is an ongoing information lifecycle need of an organization. DM began with the electronics era of data processing. But the methods have roots in accounting, statistics, logistical planning and other disciplines.

A formal document that outlines what you will do with your data during and after you complete your research is data management plan. It is also known as data sharing plan, data archiving plan or technical plan. While formulating a plan, multiple components are to be considered, including how to archive, where to and how to share data. The large volume of data both structured and unstructured are to be handled. To make the data useful, organizations are able to harness the power of the data and gain the insights.

Data Management Plan (DMP) consists of the following eight components:

1. Administrative Information
2. Data Collection
3. Documentation & Metadata
4. Ethics & Legal Compliance
5. Storage, Backup & Security
6. Selection & Preservation
7. Data Sharing, and
8. Responsibilities & Resources

However, the categorization may vary as per the context and organization also.

Research Data Management

Research Data Management (RDM) has been defined as, ''the organization of data, from its entry to the research cycle through to the dissemination and archiving of valuable results'' (Tedds and others, 2011)[12]. However, Cox and Pinfield (2014)[3] observed RDM as that which 'consists of a number of different activities and processes associated with the data lifecycle, involving the design and creation of data, storage, security, preservation, retrieval, sharing, and reuse, all taking into account technical capabilities, ethical considerations, legal issues and governance frameworks'. A highly complex set of activities in the RDM involves an array of technical challenges as well as a large number of cultural, managerial, legal and policy issues. RDM has been discussed in the literature of the library's environment for the past 5 to 6 years.

Research Data Management (RDM) aims to make the research process as efficient as possible, and meet expectations and requirements of the university, research funders and legislation. It concerns on how to 1) Create data and plan for its use 2) Organize, structure and name data 3) Keep it — make it secure, provide access, store and back it up, and 4) Find information resources and share with collaborators and more broadly, publish and get cited. Whyte A and Tedds J (2011)[16] emphasized that RDM aims to ensure reliable verification of results, and permits new and innovative research built on existing information.

Literature Review

The role of libraries in data management was discussed by many researchers. Data management involves a number of stakeholders, both within and outside the university community. Flores et al. (2015)[5] argued that the academic library is important because it occupies "the unique position, as both a facility with staff that has expertise in many of the issues surrounding RDM". Libraries provide information literacy and user education. It also curates and preserves information. Reinhalter and Wittman (2014)[9] emphasized data management by libraries, which are helping researchers. They also said that RDM planning is like, "Navigating this brave new world is simply an extension of the work that has been going on for decades."

Corrall, Kennan, and Afzal (2013)[2] studied the research support services in academic libraries by conducting a survey. They have provided a set of international comparisons in a study including RDM and other library research support services in Australia, New Zealand, Ireland and the UK, commenting, for example, that the UK appeared to be lagging behind Australasia in some critical areas. A quantitative study by Cox and Pinfield (2014), was complemented by a qualitative research (Cox, Pinfield & Smith, 2014[4]; Pinfield *et al.*, 2014), that have updated the situation in the UK and have illustrated the complexities involved in developing RDM programs at institutional level. A study by Tenopir *et al.* (2012)[11] found that a number of librarians did not have the skills needed for DM. It may be because that they were not given enough time by their institutions to provide data management services, and hence their libraries do not consider data management services a priority. A study of science librarians at Association of Research Libraries (ARL) institutions by Antell et al. reported that "Less than a quarter of survey participants (23.2% of 175 librarians) thought they had the skills necessary to provide data management services". But some libraries—such as those at Purdue University, the Massachusetts Institute of Technology (MIT), Cornell University, and the University of California—have been successful in providing RDM services and they have shared their experiences and knowledge (Knight, 2015)[6]. There are efforts by many libraries of DM (planning) and services, as can be found with a number of case studies from a variety of institutions across the board. Xia and Wang (2014)[17] examined for competencies and responsibilities of social science data librarians using 167 job postings from International Association for Social Science Information Services & Technology website for 8 years. They identified the need for the growth in data management services in academic libraries and

found an increasing demand for "data management planning" in the job descriptions.

Research Data Services

Cox and Pinfield (2014)[8] defined the Research Data Services (RDS) as services that address the full data lifecycle, including the data management plan, digital curation (selection, preservation, maintenance, and archiving), and metadata creation and conversion. Tenopir *et al.* (2012)[13] did a survey on RDS that a library offers to researchers in relation to managing data and can include informational services (e.g., for faculty, staff, or students on data management plans, metadata standards; providing reference support, or providing web guides, etc.), as well as technical services (e.g. repository, de-accessioning or deselecting data sets from a repository, or creating metadata for data sets). The survey was to find out the types of research data services that are currently offered in each library, RDS planning stages for the future, staff capacity, the leadership devoted to research data services, and the types of staff training that are allotted for RDS. The result provided a direction for the libraries that are taking in research data services, with special emphasis on the level of involvement of libraries that address the full data life cycle (Figure 1).

Components of RDM services

Zhou (2018)[18] identified the major components of the RDM Services are as follows:

- √ *Research Data Management planning* - To reduce risks and rationally allocate resources a data management plan which should be prepared.
- √ *Metadata services* – To maintain standard, interoperability and enhance data sets for helping the researchers; metadata services are provided. The metadata standards can synthesize other standards such as EML, ISO 19115, Dryad, TEODOOR, PANCAEA, etc. for various subject areas.
- √ *Research Data Storage* – To preserve the valuable research data for the long-term is an essential part of scientific research activities. It can be used to accumulate basic data for future research. The service refers to assisting scientific research personnel in submitting, storing, backing up, and updating research data. It is a core function of researching data management services. Libraries have a rich experience in resource conservation.
- √ *Research Data Mining and Sharing* - Research data is stored for its mining and sharing services. The goal of

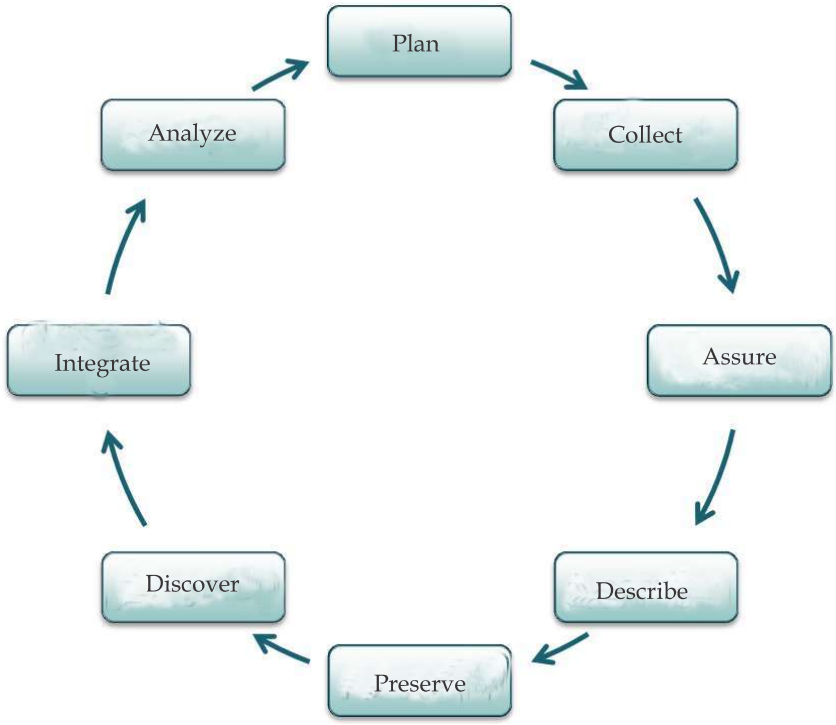

Fig. 1: Data Life Cycle

research data management is to analyze, collate and evaluate saved data, tap the correlation between these data and other related information, carry out secondary development, realize value-added data, and allow more scientific researchers to use these valuable values through certain sharing channels. Data sharing services include data publishing, discovery, retrieval, download, and other services. Data sharing permissions held with the responsibility of researchers and creators, different data and different groups of people, such as users outside the organization and users inside and outside the project, confidential data and non-confidential data, will have very different sharing rights.

√ *Research Data Reference* - Reference consulting is a kind of personalized deep service and it is also one of the core work contents of subject librarians. In order to develop reference services for research data, librarians can be embedded in the work of scientific research teams. Based on a large number of library data sources, they can actively understand the scientific research workers by focusing on data classification, uploading, storage, maintenance and sharing. The data management needs to answer the various problems they encountered in the data management process and provide them with targeted services.

√ *Research Data Literacy cultivation* - One of the necessary capabilities for academic researchers are to research and communicate. It relates to data awareness, data management knowledge and data management skills. Scientific data literacy is cyclical, emphasizing the activities of collecting, processing, evaluating, managing, and utilizing scientific data, and paying attention to the various skills required to manage data. In addition, it emphasizes the ability to analyze data, present data, and use data management tools.

In order to support effective data management and sharing, an institution needs a coherent strategy and suite of services. When delivering RDM services, the roles and responsibilities of those who may deliver and use them are given below in (Fig. 2).

Role of Library Professionals

One of the most important aspects of a librarian

Fig. 2: Components of research data management support services

is to communicate effectively with researchers for providing research data services and concepts related to data. The Data ONE Usability & Assessment Working Group adapted the data life cycle to put questions that libraries and librarians may have regarding research data services into context. A high level of variation in data curation concepts was found in the study conducted by the Purdue University Libraries or even within, fields of study, which served as a barrier between librarians and researchers (Carlson, 2012)[1]. The study resulted in the Data Curation Profiles (DCP) Toolkit (http://datacurationprofiles.org), a semi-structured data reference interview instrument that is designed to help librarians connect with researchers and identify their data needs.

Academic libraries and archives already have experienced a different array of file types: audio, images, software code, and datasets. The next step is creating data repositories for improving digital infrastructure for research data management. The main feature of data repositories is data sharing, which is increasingly recognized as integral to scientific research and publishing. One has to plan it from collection of data / metadata, interoperability, deposit in data repositories and curation. Also it should be with different collaborators who will share their data to the repositories, may conduct data interviews with scientists to assess current research lifecycle practices, document types of research data available and evaluate needs surrounding long-term management of data.

The library professional should be aware of research on various trends and practices at universities for curation of restricted research data with particular attention to the role of institutional repositories. Although as emphasized by Socha (2013)[10], some disciplines (e.g., astronomy) have a long established practice of sharing and citing scientific data sets, a very large number of researchers are still very reluctant to do so.

Capability of the librarian in handling Data Services

Does the Librarian possesses the capability for the following:

- To do Research Data Services as priority
- Level of participation with data
- Role of librarian in discovering data
- Understanding the file formats
- Level of involvement with metadata
- Stewardship role (select & deselect) – any repository that accepts data
- Capability for preservation
- Selection of repository (either subject or general)
- Role in partnering with researchers
- Data set licensing
- Level of knowledge skills
- Identification of persistent identifiers

Staff Development and skills

An international study of research data management activities, services and capabilities in higher education libraries was conducted by Cox *et al.* (2017) in Australia, Canada, Germany, Ireland, the Netherlands, New Zealand, and the UK, indicated that libraries have provided leadership in RDM, particularly in advocacy and policy development found from different countries. The library managers stressed / informed the following:

- Need for RDM development policy
- Library partnerships for collaborating (with external organizations, data centres, universities, university libraries) research data services
- Libraries and organizational restructuring
- Library RDM skill development needs for i) Data curation ii) Technical and ICT skills (data storage, infrastructure, architecture, etc.) iii) Subject and / or disciplinary knowledge iv) Knowledge of a variety of research methods (data analysis, data visualization, etc.) v) Knowledge of the research life cycle vi) Data description and documentation and vii) Legal, policy and advisory skills (intellectual property, ethics, licensing, etc.)

Pinfield and others (2014) in their article on RDM and libraries identified a number of important drivers at the institutional level. They are — i) Storage ii) Security iii) Preservation iv) Compliance v) Quality vi) Sharing and vii) Jurisdiction. They also identified the factors that are influencing the RDM developments as — a) Acceptance b) Cultures c) Demand d) Incentives e) Roles f) Governance g) Politics h) Resources i) Projects j) Skills k) Communications and l) Context.

Librarians' attitudes regarding the importance of RDS for libraries and institutions as reported by Tenopir *et al.* (2012)[11] are as follows:

- RDS are just as important as other activities
- RDS are unnecessary for librarians to provide to their users
- RDS are a priority at my library
- It will increase the visibility and impact of our institutional research
- RDS are a distraction from the library's core mission
- The library is the best suited entity to provide RDS

Conclusion

Libraries and other traditional institutions should innovate their services. Data Management provides opportunities due to the increasingly prominent value of research data and the transformation of scientific research. The complex issue of RDM involving multiple activities has been carried out by various actors since there is a need for collaborators for collecting research data which addresses a range of drivers and factors. To provide RDS in the future, the library professionals believe that they have the knowledge and skills required. They are also motivated by professional interests, patron demand and job responsibilities.

However, RDM service as a whole is still in its infancy. It needs to go through such links as policy formulation, infrastructure construction, service content design, service team formation, service user mining and service fund raising. It needs to establish a continuous and effective research data service model to promote the further development of the open access movement. Universities and their libraries need to have a deep understanding of the operational processes, best practices and influencing factors of each link and in combination with their own development.

Library professionals believe that they should have mission to RDS which are becoming an important service for academic research. They also have the opportunities to lead, engage or collaborate in workshops, seminars and publications due to the involvement of research data management services. Also, these librarians believe that research data services will increase the visibility and impact of institutional research. For increased alignment between library services and the university research mission, the library-based RDS are important opportunities.

References

1. Carlson Jake. Demystifying the Data Interview: Developing a Foundation for Reference Librarians to Talk with Researchers about Their Data. Reference Services Review. 2012;40(1):7–23.
2. Corrall S, Kennan M, Afzal W. Bibliometrics and research data management services: emerging trends in library support for research. Library Trends. 2013;61:636–674.
3. Cox AM, Pinfield S. Research data management and libraries: Current activities and future priorities. Journal of Librarianship and Information Science. 2014;46:299–316.
4. Cox AM, Pinfield S, Smith J. Moving a brick building: UK libraries coping with research data management as a "wicked" problem. Journal of Librarianship and Information Science. 2016;48:3–17.
5. Flores, Jodi Reeves, Jason J Brodeur, *et al*. Libraries and the Research Data Management Landscape, The Process of Discovery: the CLIR Postdoctoral Fellowship Program and the Future of the Academy. 2015;82–102.
6. Knight Gareth. Building a Research Data Management Service for the London School of Hygiene & Tropical Medicine, Program: Electronic Library and Information Systems. 2015;49(4):424–39.
7. National Research Council. Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age. 2009. Available from http://www.nap.edu/catalog.php?record_id=12615. [Accessed May 2019].
8. Pinfield S, Cox AM, Smith J. Research data management and libraries: Relationships, activities, drivers and influences. 2014. Available from http://doi.org/10.1371/journal.pone.0114734. [Accessed May 2019].
9. Reinhalter, Lauren, Rachel J Wittman. The Library: Big Data's Boomtown. Serials Librarian. 2014;67 (4):368.
10. Socha YM. Out of cite, out of mind: The current state of practice, policy, and technology for the citation of data. 2013. Available from https://www.jstage.jst.go.jp/article/ dsj/12/0/12_OSOM13-043/_pdf [Accessed May 2019].
11. Tenopir *et al.* Data Sharing by Scientists: Practices and Perceptions. PLoS One. 2011;6(6):e21101.
12. Tedds J. Making the case for research data management. 2011. [online] Edinburgh: Digital Curation Centre website. Available from http://www.dcc.ac.uk/webfm_send/487 [Accessed May 2019].
13. Tenopir, Sandusky, Allard, *et al*. Academic Librarians and Research Data Services — Current Practices and the Plans for the Future. [online] Association of College and Research Libraries

website. 2012. Available from www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf [Accessed May 2019].

14. Tony Hey, Stewart Tansley, Kristin Tolle (Eds). The Fourth Paradigm: Data-Intensive Scientific Discovery. Redmond, WA: Microsoft Research; 2009.
15. Wallis, JC, Borgman CL, Mayernik MS, *et al*.Moving archival practices upstream: An exploration of the life cycle of ecological sensing data in collaborative field research. International Journal of Digital Curation. 2008;3(1).
16. Whyte A, Tedds J. Making the Case for Research Data Management. DCC Briefing Papers. Edinburgh: Digital Curation Centre; 2011.
17. Xia, Jingfeng, Wang Minglu. Competencies and Responsibilities of Social Science Data Librarians: An Analysis of Job Descriptions. College & Research Libraries. 2015;75(3):362–88.
18. Zhou Q. Academic Libraries in Research Data Management Service: Perceptions and Practices. Open Access Library Journal. 2018. Available from https://doi.org/10.4236/oalib.1104693 [Accessed May 2019].