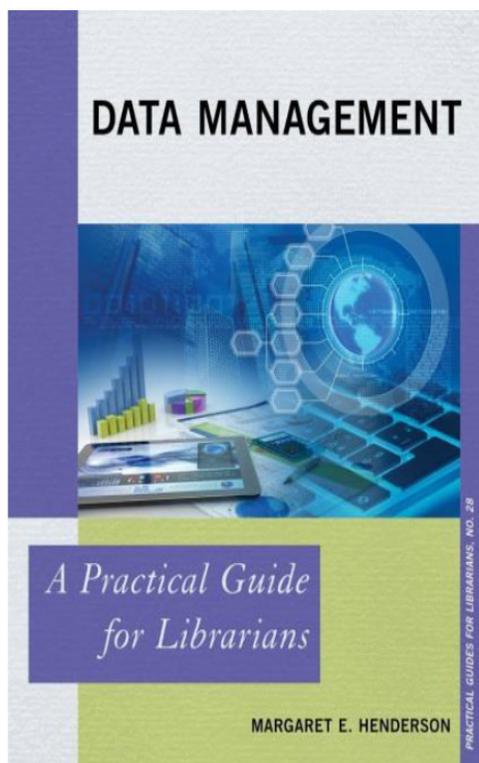


# Data Management: A Practical Guide for Librarians



## Data Management: A Practical Guide for Librarians,

By  
**Margaret E. Henderson;** Rowman & Littlefield Publishers, Lanham, Maryland, 2016, Paperback, ISBN: 9781442264380.

Research organization produces a large amount of research data annually e.g., European Organization for Nuclear Research (CERN) in Geneva is expected to produce 15 petabytes of data annually.<sup>[1]</sup> (Proyor, 2012). Therefore, it is imperative to manage research data to derive maximum benefit out of it. A study conducted in the United States estimated that Government spending on Human Genome project was \$13 billion. However, its successors have generated benefits of nearly \$1 trillion out of it. There are several ways of data management for best impact. Libraries being neutral institutions deliver services to all and are the most preferred places to initiate and manage research data services.

In her book, *Data Management: A Practical Guide for Librarians*, Henderson explains various aspects of data management and the role of library professionals. The book, with 14 chapters, is a comprehensive source of information for data librarians, computer scientists, library professionals and policy makers to understand the nuances of research data management practices prevalent in academic and research institutions. The book starts with an introductory note on what is data and why librarians should be involved in the process of research data management in academic and research institutions. Demand has been raised by various groups to make Government data open access. The author argues that library professionals are most suitable to manage research data and libraries are neutral collaborators in any organization. Librarians have always been involved in collection, storage, organization and retrieval, and dissemination of documents and the same process can be followed in data management. The author lucidly defines the role of data management in the research process and researcher

### Copyright

© The Author(s). 2018 This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

work cycle. Further, the importance of research data management and allied technologies in 21<sup>st</sup> century research, and the role of librarians are dealt with in detail. It also explains some best practices in data management in academic and research institutions, along with individual and institutional reasons to start data services in an organization. The author deliberates upon planning, collecting, describing, publishing, sharing and preserving research data for the benefit of readers. Data interview skills required are also discussed in the book so that data librarians can successfully interview researchers to help in the data collection process. Storing, curating and preserving of data are the most difficult part in research data management.

The author deals with each aspect in detail so that data librarians can efficiently handle inhibitions associated with these issues. Furthermore, it is stressed that responsible librarians need to be abreast of changes with technology, and policies associated with research data management. Henderson opines that librarians are the best choice in an organization to handle documentation of research data. The librarian's knowledge about ontologies and subject headings can support data organization and s/he is well placed to provide research data support to researchers in an organization. It is essential to use naming conventions and standardized formats, and to itemize these for researchers. Ethics of documentation have been discussed briefly in the book. There is mention of the importance of good record keeping as being crucial in data sharing and collaboration. Good record keeping leads to transparency and reproducibility of research. Basic information about the *readme* file and what it should contain are also mentioned. Types of metadata and Dublin Core metadata elements are discussed with various examples. Author stresses that prior to the start of data management, researchers should study the organizational scheme and carefully document the data in their project. It will help them to preserve and share the research data in future. Author advocates in her book that data ought to be used as scholarly output in citations, promotion and tenure renewal. Along with several benefits, data sharing also has several barriers including, lack of opportunity, human subject privacy, other privacy, and lack of knowledge about repositories. In addition, preparation of data is also identified as one of inhibits in sharing datasets.<sup>[2]</sup> Nevertheless, author is hopeful that all these barriers can be removed by the librarians with the help of data management services. Data librarians can play a vital role to anonymize the dataset to counter privacy issues raised by researchers in sharing of data. The book highlights the issues in getting credit for sharing research datasets. Thus, standards of datasets have been developed to allow citations of data. Consequently, several journals are allowing citations of data in research papers. The book covers various aspects of depositing data for public access and stresses that subject-specific data repositories should be preferred over general repositories while depositing research data. Repository librarians familiar

with the documentation have a vital role to play in educating researchers about metadata standards and procedures involved in depositing data.

The book also explains the reasons to make a data management plan, how to collect information for a data management plan and writing data management plan for a grant. Further, the book also describes the advantages of data management planning. Several templates are given for writing a data management plan. In addition, multi-institution template projects are also given in the book wherein the author gives step-by-step procedure to conduct interviews viz., explaining the data to be collected, methodologies for data collection and management, ethics and intellectual property issues, plan for data sharing and preservation issues.<sup>[3]</sup> The book discusses the various ways of starting data management services in academic and research institutions. It has a chapter explaining data management services wherein environment assessment, librarians' skills assessment, planning services, organization of services, potential RDM services and evaluation of research data management (RDM) services etc. have been discussed in detail. The chapter concludes with remark that evaluation of RDM services should be communicated to library head and institutional stakeholders. Besides this, it is most important to include not only quantitative data but also impact of the services on individual stakeholders.<sup>[4]</sup> Author also advocates a logic model to develop an evaluation method covering all aspects such as, resources, activities, outputs and outcomes. The staffing pattern required for a research data service is laid out and it is stressed that library and non-library skills of librarians should be considered while staffing RDM services in an institution. The book also covers the partnerships required for successful management of research data. It is stated that interactions with the academic community of the institution is most required and liaison with persons having significant connections in the organization could be crucial for successful RDM services. Metadata experts in the technical section of the library have to learn additional skills for adding metadata to research data. It is recommended that outreach programmes be organized for librarians who are involved in data management because they can refer people to the service when needed and they can mention the service while giving orientation talks. Author concludes the chapter highlighting that RDM team members should learn leadership, not management, and RDM team should treat collaborators as trust associates, not subordinates. Author also explains the data life cycle steps and expanded services associated with them. Following are the data life cycle steps explained: (a) plan, (b) collect, (c) describe, (d) process and analyze, (e) publish and share, (f) preserve, (g) reuse, (h) disciplinary data, and (i) specialized data services. Author stresses that interface of the data repository must be user-friendly so that users can deposit data easily; librarians should be prepared so that they can help others to enter metadata associated with datasets. Further, in a subse-

quent chapter, the author states that data information literacy competencies can help in developing a framework for different types of data management instructions.

Therefore, RDM services should provide training and support for research data collection, management and analysis, and in sharing to research scholars and faculty members in the organization from time to time. The dedicated chapter on teaching data also discusses library instruction formats, learning paradigms, instructional design models, lesson planning, motivation, data instruction, course integrated instruction and curriculum-integrated instruction. A separate chapter deals with reusing data which discusses the health and climate of data reuse. In addition, skills and tools required for data reuse are also explained briefly. The author has listed several general repositories and visualization tools for the benefit of the readers. Last chapter of the book covers the role of librarians in data management. Data librarians have an opportunity to highlight copyright issues associated with data ownership. Researchers ought not to go against institutional policies. Data librarians can plan sustainable work to setup the workflow, educate research scholars, create data dictionaries, and capture code for analysis. Overall, the book is a comprehensive source of information for library professionals, computer scientists, LIS students and policy makers, to understand the nuances

of research data management. The concepts relating to RDM have been discussed in detail. The language of the book is easy to understand, and various tools and repositories have been provided for those who intend to start data services in their organization. Undoubtedly, the book is not only an invaluable guide for staying update and managing the data in academic and research institution, but also as a tool of data librarianship into the broader context.

## REFERENCES

1. Britt R. (2010), NDF Science Resources Statistics. National Science Foundation, available at: <https://www.nsf.gov/statistics/>, accessed on June 1, 2018.
2. Federer *et al.* Bio-medical data sharing and reuse: Attitude and Practices of Clinical and Scientific Research Staff. Plos One. 2015;10(1/2). e-01129506, available at doi:10.1371/journal.pone.0118053.
3. Research Data Oxford (2016), Data Management Planning, accessible at: <http://researchdata.ox.ac.uk/home/managing-your-data-at-oxford/data-management-planning/>, accessed on May 27, 2018.
4. Albert AB. Communicating library value –the missing piece of the assessment puzzle. Journal of Academic Librarianship. 2014;40(6):634-7.

**Reviewed by**

**Raj Kumar Bhardwaj**

Librarian, St Stephen's College,  
University of Delhi, Delhi-110007, INDIA.

Email: [raajchd@gmail.com](mailto:raajchd@gmail.com)

DOI: 10.5530/jscires.7.2.24