

the local level, it is indispensable to have a book focused on this sector of archives.

© Dennis Roman Riley
New York State Archives

NOTES

¹ For more on the status of the SAA-ACRL/RBMS Joint Task Force on Public Services Metrics, see <https://www2.archivists.org/groups/saa-acrlrbms-joint-task-force-on-public-services-metrics>.

² “Best Practices for Volunteers in Archives,” Society of American Archivists, August 2014, <https://www2.archivists.org/standards/best-practices-for-volunteers-in-archives>.

The Data Librarian’s Handbook

By Robin Rice and John Southall. London: Facet Publishing, 2016. 192 pp. Softcover and EPUB. \$95.00U.S., £54.95U.K. Softcover ISBN 978-1-78330-047-1; EPUB ISBN 978-1-78330-183-6.

In *The Data Librarian’s Handbook*, Robin Rice and John Southall examine the role of the data librarian in academia. Detailing the foundation and development of the field of data librarianship, this work is a valuable resource to new data librarians or curators, or anyone interested in the field. Authors Robin Rice and John Southall, data librarians at the University of Edinburgh and the University of Oxford respectively, have provided a comprehensive outline of the field of data librarianship and illustrate how data librarians fit into the overall research and education experience.

The first chapter, “Data Librarianship: Responding to Research Innovation,” provides an overview of data librarianship and the phenomenon of the “accidental data librarian,” comparing the similarities and differences between data librarianship and “traditional librarianship.” Data librarianship and traditional librarianship share the same basic goals “to support learning and the spread of knowledge” (p. 1). Data librarianship, however, is concerned with a specific type of information—digital data. Likewise, the distinctions between archivists and data librarians are very subtle and vary based on the environment and institutions in which they operate. While both positions focus on long-term preservation, archivists manage records in a variety of formats, including born-digital materials. Data librarians, on the other hand, are more likely to handle born-digital objects, but can also manage analog data that has been digitized. In addition to describing the “accidental data librarian,” the authors provide an overview of outside influences that encourage the management of

data, such as governmental and funding agency requirements, and the development of institutional repositories.

Chapter 2, “What Is Different about Data?,” addresses perceptions of data, different data types, and the importance of data management planning across disciplines. Some of the main concerns of data librarianship include data creation, reuse, and intellectual property. This chapter also discusses metadata, data in contrast to “big data,” and the importance of advocacy. After describing what data is and how it should be managed, the authors detail the skills that can be utilized to encourage collaboration and data management.

Rice and Southall illustrate the ways in which data citation, data discovery, and reference interviews are opportunities to promote preservation and access of data and educate stakeholders of best practices, while encouraging open sharing. Chapter 3, “Supporting Data Literacy,” reinforces the importance of sharpening skills and staying up to date on the latest strategies and techniques to promote data literacy, and learn more about data management or the creation and development of policies and procedures for effective data curation. The policies and procedures created may inform actions of stakeholders. The data librarian is in a position to encourage best practices in information literacy and data management by remaining diligent and informed and effectively disseminating relevant information.

What does it take to build a data management service? Chapter 4, “Building a Data Collection,” explores the development of policies for curating and preserving data. The authors provide recommendations for developing a data management policy covering all the aspects of the data life cycle from acquisition and providing access, to promoting and sustaining collections. Rice and Southall discuss embedding data and digital collections within library collections and making the management of digital collections a part of normal library functions.

“Research Data Management Service and Policy: Working Across Your Institution,” chapter 5, is geared toward research data management plans and their implementation. In this chapter, the authors discuss why a data management planning policy is necessary and how to develop one. Rice and Southall provide an overview of how librarians may fit into this process, encourage progress, and become a valuable resource to researchers and other stakeholders while building alliances.

In chapter 6, “Data Management Plans as a Calling Card,” the authors provide eight case studies from various disciplines detailing their data management plan initiatives and experiences. Each example presents a unique perspective along with tips and advice for data librarians. For example, in a vignette from the London School of Economics and Political Science, applicants were given access to data management plans from successfully funded projects to

use and adapt for their own projects. This service attracted more users and led to more interest and referrals from researchers for assistance with data management planning (p. 88). These case studies support the argument that data management plans may be the most powerful and effective strategy to build alliances and support within organizations and strengthen research and scholarship. Data management plans are often required by funding agencies, therefore forcing researchers to give forethought to their data. Because these requirements can be difficult and time consuming, many researchers are happy to accept assistance creating them. This becomes an opportunity for data librarians to provide guidance to researchers and display their expertise in these areas while building relationships between themselves and researchers. With a better understanding of what services are needed and how data management plans are created, one can consider where the data will be preserved and made accessible.

Chapter 7, “Essentials of Data Repositories,” is a discussion of data repositories—how to develop, manage, and promote the use of and participation in repositories. The authors discuss the perceived differences in the terms “repository” and “archive.” Some view archives as simply long-term storage, while the term “repository” has a broader meaning that departs from traditional views of archival storage (p. 103). This chapter focuses mainly on the data curation process associated with data repositories, providing insight on platforms, metadata, and types of data. The authors include a description of all the necessary elements of a data repository and stress the importance of interoperability.

The data held in these repositories often have unique considerations that must be addressed, such as privacy. In chapter 8, “Dealing with Sensitive Data,” Rice and Southall describe how to deal with sensitive and confidential data and how to work with researchers who may be wary of giving access to such data. The authors discuss providing the proper authorizations for access and taking the due diligence to protect sensitive data. This chapter allows the data librarian to view data from the perspective of researchers/creators and users, and helps to find a balance that promotes reuse. Each data set has its own unique qualities, and within academia there is even more variety in research and data types across disciplines.

Data sharing in academia within the social sciences, the sciences, the arts, and the humanities is explored in chapter 9, “Data Sharing in the Disciplines.” Rice and Southall provide examples of the changes and unique attitudes toward data in disciplines such as psychology, astronomy, climatology, and genetics. The authors identify changing approaches to data reuse and how data librarians can meet these needs. For example, in the social sciences, data reuse for smaller scale projects has been uncommon, but researchers are being encouraged to share their data regardless of scale and to value replicability and verification (p.

138). Outside of academia, citizen scientists who collect data for large projects will likely share them without the same reservations professional researchers may have. Physicists, for instance, may have more concerns when sharing data because the data may not be worth the time and money, or because they have a publication bias, only publishing and sharing significant or positive results (pp. 140–41).

Chapter 10, “Supporting Open Scholarship and Open Science,” discusses the “data revolution” and the emphasis on open science and open access. Rice and Southall follow the evolution of the open access and free software movements and the ways in which research methods evolved to embrace open access and reinforced the importance of data sharing, and the technology-driven acceleration of data preservation, use, and reuse. These changes require researchers to receive credit for the data they produce while allowing others to access the data and the idea of making data a “first-class research object” (p. 153). The authors discuss the relationship between big data, the “data revolution,” and the scientific paradigm. Big data has changed how research is conducted, leading researchers to rethink how they process their data and create and use new tools. This chapter also addresses the importance of reusing data as quality control, allowing other researchers to test reproducibility and helping science to self-correct.

This work speaks to the complexity of the roles and concerns of data librarians. When reviewing this work, I prepared myself for a slow, technical, dry manual, but I was surprised to find this book to be well written, easy-to-read, and enjoyable. Although this volume is geared toward data librarians, it is written in a way accessible to those outside the field because the authors avoid field-specific jargon and focus on practical and relatable topics relevant to anyone who manages data. One of its most useful features is the “Key take-away points” and the “Reflective questions” at the end of each chapter. These are useful for their summaries and questions that encourage readers to reflect on the chapter’s content and how the information it presents is relevant.

The authors address many issues that I confront in my own position as a data curation librarian, and the advice and guidance they give reflect many of the decisions we have made within my own institution to build successful data management services and support. In general, I agree with their assessments of the particular skills that make data librarians a valuable asset to researchers and of the challenges, the roadblocks, and the possible solutions. Despite the limitations many librarians face, this volume provides relevant and detailed case studies and best practices that will make the task much more manageable.

From data management plans to the research life cycle, *The Data Librarian’s Handbook* guides the reader through the roles, responsibilities, and challenges

of data librarianship. For currently practicing or future archivists and librarians interested in the quickly evolving field of data management and curation, this book is an ideal resource, providing insight and a foundational understanding of data management and data in their various forms.

© **Kayla Siddell**
Indiana State University