Journal of Information Literacy

Book Review

Book review of Koltay, T., 2021. *Research data management and data literacies*. San Diego: Elsevier Science & Technology. 2021. 196pp. 9780128244753. £45.71. Pbk.

Hozefa Ramgadwala, MLIS student, UCL. Educator & Academic Liaison Librarian, Aljamea-tus-Saifiyah. Email: Huzaifa.ramghad@jameasaifiyah.edu. ORCID: 0000-0003-3343-1002. Twitter: @HRamgadwala.

In this book the author provides a fresh perspective for managing research data by relating it with data literacies (in the plural), acknowledging the fact that multiple literacies are involved in managing data, while still keeping the focus on information literacy (IL): he considers this monograph an extension on his previous book *Research 2.0 and the future of information literacies*. This book will be of special interest to researchers and students in higher education (HE) and it targets information professionals, particularly academic librarians, who are aiming to support researchers and HE students. In the introduction the author emphatically focuses on the necessity of data literate students and professionals for the management of research data and emphasizes without them research data management (RDM) cannot be successful, or at least may prove ineffective (p. x).

The book could be divided into four sections: 'chapters discussing data and information' (Chs. 1 & 2), 'discussion about the most essential ingredient — quality of the data' (Ch. 3), 'Research Data Management' (Ch. 4), and 'chapters explaining new literacies in the information ecosystem and the role of information professionals in it' (Chs. 5 & 6).

In the introduction the author outlines how data was previously only seen as a by-product of businesses and had little value as an asset, but that recently, due to data mining and machine learning, data has pervaded every sector of society (p. 1). The author then describes the importance and relevance of data sharing and reuse: he asserts that 'the commitment to making knowledge more easily accessible to others arguably goes back to the advent of scientific journals' (p. 21). He explains that free accessibility of research data, 'which was collected with great effort and in lengthy processes' (p. 21) will provide support and prove financially efficient for further research. He identifies another relevant factor: misinterpretation of data, which can be a byproduct of free research data. He also describes how coauthoring publications enhances their value (p. 37) and on the other hand how information overload (IO) can cause difficulties in managing research data (p. 38).

The author always provides both sides of the argument without being biased towards one perspective. For example, while discussing the terms 'data' and 'information', he explains the difference between them as well as how they are used interchangeably. Similarly, when professing the benefits of data sharing, he also discusses the pain points of data misinterpretation and misuse.

In the next section, the author discusses an essential issue regarding research — data quality. He also provides a few pointers for increasing the level of data quality, such as understanding researchers' problems, improving the peer-review process, and increasing incentives for open research (p. 52) and believes that the fate of data quality depends upon these three sets of data professionals: data managers, data librarians, and data scientists (p. 53). The author is keen on understanding one particular quality of data — 'trustworthiness'. He dedicates an entire section

of the chapter to explaining how trust is a distinguished data attribute and how it relates to 13 different elements, such as origin, processing, and authenticity of the data (p. 59). He further discusses the management of big data and how often the incomplete and uncertain nature of big data may negatively influence its quality (p. 63).

An interesting feature about this author is that he uses metaphors and examples to make his point. For instance, while explaining the significance of data governance, he compares it with the ancient Indian fable of an elephant in a dark room, which can be perceived differently depending upon where you touch it: the tail as a snake, the leg as a tree. Similarly, Koltay says, cross-functional perspectives on data governance vary and this should be taken into consideration when relating it with other literacies (pp. 64–65).

The next section focuses on one of the key elements of the book — Research Data Management. The author starts the chapter with an interesting debate about how research data differs from regular data. If one argues that research data is the data specifically collected as part of research, in contrast to data that is curated or preserved, then a question emerges: can the curated data be used in research, or can research data be curated at a later stage? The author further explains the ambiguity through terms such as RDM (Research Data Management) and RDS (Research Data Services) as there are significant similarities in these two service activities (p. 77). Further, he discusses the role of academic librarians in providing RDM services and the skills and competencies required for them to provide this service efficiently, which includes identifying and locating datasets, promoting data sharing and reuse, and teaching data literacy to researchers and students (pp. 81–82). He also provides significant literature supporting the idea of allocating responsible human resources to lead RDM services. These services can be broadly divided into three types: Education, Expertise, and Curation (p. 89).

In the final chapters, the author explains the emergence of 'new literacies', broadly meaning the development of new social practices with new technologies (p. 109) and the roles of LIS (Library & Information Science) professionals and DS (Data Science). Further, he delves into the connection and dependency between IL and data literacy and how the latter has emerged from the former (p. 110). He explains how the skills needed for successfully managing research data are tied to both IL and data literacy (p. 113). Data literacy, as understood by ACRL, focuses on understanding how to find and evaluate a particular version of a dataset and its origin (the person responsible for it), and how to use the data ethically (p. 120). He then provides a framework for data literacy and briefly discusses other relevant literacies such as visual literacy and artificial intelligence. The author also represents the convergence of Information Science, Data Science and e-Science in a graphical presentation, depicting the interconnection between these roles in an academic context.

In conclusion, this book can be a complex read for anyone not familiar with data terminologies and library jargon. However, given the way the author has provided examples and illustrations and interlinked the chapters (by providing references to previous chapters), this book offers an interesting gateway for a beginner as well as a professional.

This book would be a valuable resource for any researcher or HE student; similarly, it would stand as a treasured source for an academic librarian or any information or data professional.

.