Book Review

Timothy Michael Mulcahy¹

Louise Corti, Veerle Van den Eynden, Libby Bishop, and Mathew Woollard (eds). *Managing and Sharing Research Data: A Guide to Good Practice*. 2014. Los Angeles, London, New Delhi, Singapore, Washington, DC: SAGE Publications. 222 pp., ISBN 978-1-4462-67264, £25.99.

Rapid technological changes are reshaping the way we discover, access, and manipulate data and are spawning new and innovative processes that facilitate the effective management and exchange of information between agencies, independent of their geographic location or proprietary infrastructure (Heus and Gregory 2010). Increasingly, however, these changes call into question traditional statistical data management approaches and systems. As government and private-sector data producers move away from traditional survey approaches (for example, ‘design-collect-analyze-publish’) toward more integrated structures, the challenge for producers of statistics is to exploit recent technological advances to develop new and innovative ways to manage, access, process, discover, and visualize data (Lorenc et al. 2013). Old paradigms need to be revisited, prior assumptions reviewed, and traditional research methods re-evaluated and updated as appropriate.

Corti et al.’s (2014) book, *Managing and Sharing Research Data: A Guide to Good Practice*, addresses these daunting challenges. This clearly written, easy-to-follow handbook highlights best practices in research data management and sharing. However, it is not simply an anthology of data management and data sharing best practices, it is a comprehensive and contemporary primer on nearly every aspect of the research process firmly rooted in the research data lifecycle (Humphrey 2006).

The book is organized into eleven related chapters and applies to a wide range of audiences, covering topics such as management and sharing of data, the research data lifecycle, data management planning, documenting and providing context for data, formatting and organizing data, storing and transferring data, legal and ethical issues, rights relating to research data, strategies for collaborative research, secondary research, publishing, and citing research data. The book is particularly timely in that funding agencies around the globe are formulating new laws, policies, and procedures that strongly encourage, if not require, researchers and grantees to provide data management plans or strategies as an integrated component of their research proposals. While these efforts are sound and aim to encourage knowledge sharing, collaboration, and open access to publicly-funded research data, they also raise new questions and challenges for data producers and end users that are adeptly addressed in this practical handbook.

¹ NORC at the University of Chicago, Bethesda, MD, U.S.A. mulcahy-tim@norc.org

In Chapter 1 the authors introduce a recurring theme that permeates the book: the need to ensure high quality, sustainable research in a responsible and efficient manner, and to do so in a way that meets the replication standard (King 1995) and ensures the ability to share and reuse research data in perpetuity. The authors also highlight the benefits of managing and sharing research data effectively, noting concerns voiced by some researchers, and providing tangible examples of successfully implemented data management plans. Chapter 2 notes the critical importance of properly indexing, archiving, and curating data to facilitate future uses and reuses of the data. Readers are encouraged to utilize a data lifecycle approach, one that is at once specific to the detailed nuances of each stage of the research process while also generic enough to apply more widely to disciplines outside of social science.

In Chapter 3, the authors emphasize the critical importance of early planning and formulating data management plans to help design and implement successful research efforts, ensure that adequate resources are in place, and clearly articulate, assign, and manage individual roles and responsibilities. Examples are provided from the UK and US for planning, documenting, formatting, storing, confidentiality, ethics, consent, copyright, and sharing.

Chapters 4–8 directly address the myriad challenges involved in data sharing and provide tangible strategies to overcome these limitations. Chapter 4 emphasizes that data alone are of limited usefulness and that they must be accompanied by proper documentation and context. Data are not simply collected to serve our current needs, they are meant to be preserved in perpetuity to allow future use and reuse. Chapters 5 and 6 extend this theme to the importance of formatting and organizing data for long-term use, storage, and transfer. Chapter 7 discusses the legal and ethical frameworks involved in data sharing and shares valuable insights into the process of obtaining informed consent (including unknown future uses), modern techniques for perturbing and anonymizing data for safe use, and regulating access to data. Chapter 8 focuses on rights related to research data, and provides an in-depth discussion on Intellectual Property Rights in data, including copyrights and exemptions, database rights, freedom of information, and licensing.

In Chapter 9, the authors point out that collaborative research is becoming an increasingly common phenomenon and caution that this newly emerging paradigm presents particular challenges to ensuring high-quality, sustainable, and reusable research. Although collaborative research heightens the need to develop and implement sound data management plans, the authors provide effective strategies to assist in this regard, for example, developing standard procedures, protocols, and policies, and clearly assigning roles and responsibilities across collaborating entities.

Chapter 10 reinforces the notion that archived data at some later point may become important historic research materials and discusses the opportunities and limitations to conducting secondary analyses. On the one hand, respondent burden is reduced, data linking is made possible, and new data sets and derivative products may be created. In addition, reanalyzing data facilitates reinterpretation and allows new questions to be asked of archived data. This provides a mechanism for validating and replicating prior work – a critical check and balance on policy and programming decisions derived from earlier findings. On the other hand, the authors caution that truly replicating historical research is hardly a trivial task, if possible at all, as studies rarely canvass identical social phenomena.

The final chapter discusses the nuances involved in publishing and citing research data, yet another area that has borne witness to dramatic changes in recent years. The authors conclude by providing rich examples of data centers, institutional data repositories, traditional and digital data archives, and associations that are on the cutting edge of publishing and data citing.

In the main, practitioners and students will find this book to be extremely valuable. Written by and for researchers, professional researchers and students alike would do well to keep a copy of this guide book close at hand. Each chapter is self-contained, allowing readers to peruse individual sections to learn more and/or refresh their knowledge and understanding of a particular issue or topic of interest and is accompanied by a comprehensive bibliography that serves as a useful guide for further research. Finally, it provides empirical case studies and “hands-on” exercises and activities that drive home the core concepts of the modern-day research process.

References

- Heus, P. and A. Gregory. 2010. Maximizing the Potential of Data – Modern IT Tools, Best Practices, and Metadata Standards for SBE Sciences. Open Data Foundation. Version 1.0.
- Humphrey, C. 2006. “e-Science and the Life Cycle of Research.”
- King, G. 1995. “Replication, Replication.” *Ps: Political Science and Politics* 28: 443–499. Available at: <http://j.mp/jCyfF1>.
- Lorenc, B., P. Biemer, I. Jansson, J. Eltinge, and A. Holmberg. 2013. “Prelude to the Special Issue on Systems and Architectures for High-Quality Statistics Production.” *Journal of Official Statistics* 29: 1–4. DOI: <http://dx.doi.org/10.2478/jos-2013-0012>.