Policies and Background Literature for Self-Education on Research Data Management: An Annotated Bibliography

Abigail Goben
agoben@uic.edu

Rebecca Raszewski
raszewrl@uic.edu

University of Illinois at Chicago
Chicago, Illinois

Introduction, Scope, Methods

This annotated bibliography reviews U.S. federal policies, articles, and books to assist librarians who are self-educating on research data management or are seeking background reading material to gain insight into the present landscape of data management and librarianship.

Librarians navigating research data management self-education have an increasing body of literature to choose from, which may become overwhelming. While an excellent and expansive bibliography of current research has been developed and is being maintained (Bailey 2015 Jul 1), those just starting to explore this area would benefit from a more targeted selection. This annotated bibliography, sorted by date of publication, addresses the policies shaping the data management landscape, provides an overview of articles to give context to the changes in libraries in the past decade, and identifies current self-education books that will provide more comprehensive insight on research data management, data information literacy, and data science topics.
Librarians seeking self-education resources beyond literature are referred to the webliography "Research Data Management Self-Education for Librarians: A Webliography" (also in this issue).

- **U.S. Federal Policies**
- **Articles**
- **Books**

**Section A: U.S. Federal Policies**

As U.S. federal funding agency requirements have been changing over the past decade, it is valuable for librarians to be aware of some federal documents that have influenced data retention and management.


This policy update from the National Science Foundation initiated the need for a data management plan for all future grant proposals. While some areas already had this requirement in place, this policy update made it agency-wide and prompted greatly increased interest in research data.

[http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf)

This memo from 2013 called for all large federal agencies to address the need for broader access to the research that they fund. Responses addressed both the final research products, including articles and their affiliated data, and are being released throughout 2015.


Supplemental to the OSTP memo, this memo instructed agencies to improve making information, specifically including data, interoperable, accessible, and reusable.
This table, compiled and maintained by a number of data librarians, provides a visual checklist of what will be expected in data management plans from each of the federal agencies.

**Section B: Articles**

A comprehensive review of articles on research data management and libraries would easily encompass an entire book. Selected here are a few articles to provide an overview of the trends and emerging research of the past decade and a bibliography for further reading.

http://www.dlib.org/dlib/september07/gold/09gold-pt2.html
Gold's articles outline the emerging challenges in advance of federal directives and the fascination with Big Data. This provides both an overview of areas for libraries to tackle data services including GIS and bioinformatics and looks ahead to the librarian's role with data.

http://www.ariadne.ac.uk/issue64/salo
Salo tackles the notion of shifting librarian activities from traditional duties to data services. She identifies the areas where there is likely to be disconnect from standard librarian practice and emerging needs and bluntly states the challenges ahead.

Borgman provides a detailed exploration of motivations, challenges, and incentives for research data sharing. This article serves as an excellent introductory primer, reviewing data complexity and current practices where assistance might be most useful.


Wallis, et al. report on the challenges of sharing data gathered over a decade from a specific research group. While the researchers show willingness to share and use others' data, their preference was to share with trusted colleagues, reinforcing a 'gift culture' based on networks and bartering rather than a more open sharing or selling model.


Erway, et al. details the university concerns surrounding research data with an eye on developing an institutional policy. This article identifies the various stakeholders, including the library, and reviews the various questions that a policy may need to answer. The article can be used as a way to gain ideas on institutional collaboration and to review issues that may arise.


As libraries have begun to engage in research data management activities, this article reviews the awareness and service landscape among science librarians three years after the National Science Foundation policy change. This article
provides insight into what science librarians saw as immediately important and future trends.

Targeted at scientists looking to improve their data as a scholarly object, Goodman, et al., provide very simple techniques supplemented with a short list of recommended resources that researchers can immediately implement to better facilitate data access and reuse. This article would serve as a nice introduction for researchers and students with some familiarity, but needing specific ideas on behaviors to adopt or change.

As librarians are learning about data management, it is beneficial for them to practice the skills and questions with library data. This article uses reference desk statistics as an example of a common dataset that can be used as an educational tool.

In this follow-up study, Tenopir et al seek to determine changes in the previous five years on perceptions and behaviors by scientists regarding data sharing and withholding and the factors influencing their behavior. This article shows the attitude changes and emerging concerns as data sharing has become more widely expected.

Bailey curates a selection of English language books and articles surrounding research data. This extensive list
includes subject discipline specific articles in addition to more general topics across library, computer science, digital curation, and other subject journals.

Section C: Books

A number of texts have been written recently to address the data management self-education needs of librarians. Along with providing broader context of the changing scholarly landscape, these books identify suggestions for implementing programs and several provide case study examples.


Borgman examines infrastructural needs as research and scholarly communication transitions in the early 21st century. She outlines emerging social challenges as well as policy, legal, community, and technological issues. Borgman also discusses issues with collaboration and data sharing, describing the new tools researchers are likely to need. Overall, the book serves as a call to action for librarians, information professionals, scholars, and information technology specialists and provides a framework for opportunities and challenges ahead.


This guide provides a succinct overview of data management opportunities for librarians. With an emphasis on collaboration with units beyond the library, Krier and Strasser develop a framework to allow for success of these new activities and is targeted specifically at librarians looking to create new services. Importantly, Krier and Strasser highlight the need for a team effort from the library, not just the hiring or appointing of a solo librarian to address data services.


This text offers a comprehensive insight into research data management as a library concern. Each chapter frames a
different data management challenge, such as assessing services, and provides guidance specifically intended for researchers working in a specific area. Four case studies examining successes and challenges provide a view into robust library data management programs at research institutions.

Briney, K. 2015. Data Management for Researchers: Organize, maintain and share your data for research success. Exeter, United Kingdom: Pelagic Publishing. Briney outlines practical strategies for researchers in tackling and managing research data. She guides researchers through a variety of data problems including storage, documentation, organization, and preservation, clearly explaining pitfalls and best practices. While written for the individual researcher, it provides a detailed roadmap for librarians tackling services beyond data management plans.

Carlson, J. & Johnston, L.R., editors. 2015. Data Information Literacy: Librarians, Data and the Education of a New Generation of Researchers. West Lafayette, IN: Purdue University Press. http://www.datainfolit.org Going beyond the self-education for librarians and on to developing data information literacy for students and research faculty, the authors provide extensive case studies that explore different techniques for developing and assessing data information literacy in the classroom. This book provides a framework for establishing and evaluating instruction to meet an emergent educational need. Ongoing additions to this project are available at the project web site.

Afflet, A. 2015. The Accidental Data Scientist: Big Data Applications and Opportunities for Librarians and Information Professionals. Medford, NJ: Information Today, Inc. Affelt tackles data science terms and frameworks, providing real world examples and ideas for where librarians can evolve into this new field with skill adaption and education, instead of focusing solely on research data management. This text transitions from the internal library focus to the broader area of data science.

Schutt and O'Neil give an overview of data science and Big Data, algorithms, data visualization, data journalism, and data engineering. Targeted as a reference text for those going into data science, this is particularly recommended for those interested in further exploring the field and goes far beyond research data management.


Borgman explores the current landscape of challenges for data. She details the present issues including the inaccessibility of data, scholars' unwillingness to share, and datasets that are unusable even if shared. She also reviews how having a large volume of data does not sort out the problem of not having the correct data. Borgman provides excellent case studies detailing what data looks like in sciences, social sciences, and the humanities. Finally, Borgman tackles the infrastructure issues including policy, citation, and discovery of data.

References