

# Research Data Management Self-Education for Librarians: A Webliography

Abigail Goben

[agoben@uic.edu](mailto:agoben@uic.edu)

Rebecca Raszewski

[raszewr1@uic.edu](mailto:raszewr1@uic.edu)

University of Illinois at Chicago  
Chicago, Illinois

## Introduction, Scope, Methods

As data as a scholarly object continues to grow in importance in the research community, librarians are undertaking increasing responsibilities regarding data management and curation. New library initiatives include assisting researchers in finding data sets for reuse; locating and hosting repositories for required archiving; consultations on workflow, data management plans, and best practices; responding to changing funder policies ([Whitmire, et al. 2015](#)) and development of department or institutional policies. Librarians looking to provide services or expand into these areas will need both foundational resources and information about engaging the network of librarians exploring data. This webliography is intended for librarians seeking to enhance their own knowledge and assist peers in improving their data management awareness.

*ISTL* published a [webliography on digital research data curation](#) in 2010 ([Westra 2014](#)); however, the research data landscape has greatly changed since then. Rather than limiting the scope to the natural sciences, this webliography examines data management across disciplinary boundaries, though still primarily focusing on United States STEM research data.

The selected resources were chosen for broad appeal across the spectra of librarians engaging with research data from novice to expert. Other factors considered included endurance of the resource over time and the authority of the materials. The webliography includes both freely available materials and educational opportunities that have costs affiliated with them, such as courses offered by professional organizations. For background literature and policy information, librarians should refer to Policies and Background Literature for "[Self-Education on Research Data Management: An Annotated Bibliography](#)" (also in this issue).

The webliography is organized by content type, first with more foundational materials such as established data management curricula and then with current awareness and community materials such as social media.

### [Foundational](#)

#### [Full Curricula](#)

[School of Library and Information Science Certifications](#)  
[Online Continuing Education and Massive Open Online Courses \(MOOCs\)](#)

### [Current Awareness and Networks](#)

#### [Journals](#)

#### [Data Sources](#)

#### [Library Professional Organizations](#)

#### [Data Professional Organizations](#)

#### [Hands-On Tools](#)

#### [Social Media](#)

## **Foundational**

Foundational materials are defined in two ways: broad material about data management and general skill sets. These resources are intended to give an oversight of multiple aspects of research data management.

### **Full Curricula**

Several curricula were developed as educational tools both for librarians seeking self education or lesson plans to use with researchers after the release of the National Science Foundation Data Management Plan Requirements ([National Science Foundation 2010](#)). These resources serve as an introduction to the data life cycle, data management plans, and best practices. A point to remember is that these curricula were developed at an early period and there have been significant changes in funding agency and institutional policies in the interim. While some continue to be updated

with case studies and additional material, their efficacy is slightly hampered in this respect.

### **California Digital Library DMPTools Webinars**

[https://dmptool.org/community\\_resources](https://dmptool.org/community_resources)

The California Digital Library is best known for hosting the DMPTool software, which guides researchers through the creation of a data management plan. Their educational materials also include an extensive introductory webinar series which explore topics related to data management planning and also implementing the DMPTool and services at your institution. The webinars and slide decks are available under a Creative Commons Attribution License. Updates to the DMPTool software are ongoing.

### **Data Observation Network for Earth (DataONE) Education Modules**

<https://www.dataone.org/education>

This curriculum provides a broad overview of research data management activities, such as data quality control and data citation. The DataONE project was originally funded by the National Science Foundation and focuses on environmental science. Ongoing webinars are archived that offer updates to the material. The education modules include a series of MS PowerPoint files with extensive notes that can be presented by themselves or modified for workshops. All of these modules are available under a Creative Commons Zero license - No rights reserved, though citation and attribute is requested.

### **MANTRA: Research Data Management Training**

<http://datalib.edina.ac.uk/mantra/>

Launched and hosted by the University of Edinburgh, this curriculum was funded by JISC in the UK and includes a DIY kit specifically designed for self study for librarians. The materials are available for remix and reuse under a Creative Commons Attribution license and include online units and offline tutorials. The curriculum continues to be updated by the librarians at the University of Edinburgh; the most recent release of MANTRA, the fourth edition, was in September 2014.

## **New England Collaborative Data Management Curriculum (NECDMC)**

<http://library.umassmed.edu/necdmc>

This resource provides an overview of the data lifecycle and takes the learner through challenges of research data management including storage, legal and ethical considerations, reuse, etc. in the form of seven modules with Microsoft Word and PowerPoint resources. Developed at the Lamar Soutter Library at the University of Massachusetts Medical School in collaboration with a group of libraries, this was funded by a grant from the National Library of Medicine and launched in Spring 2013. The modules and activities are available for download and reuse under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 United States License.

## **School of Library and Information Science Certificates/Specializations**

The following library schools offer fully online opportunities for librarians to earn degrees or certificates in data curation, or data science. If you are interested in only taking a single course, it is recommended that you contact a library school directly to determine their present offerings.

### **Indiana University Online Certificate in Data Science**

<http://www.soic.indiana.edu/graduate/degrees/data-science/online-certificate/index.html>

Four courses must be taken from a list of approved course offerings to earn the certificate. It is expected that attendees have some programming language experience and some database development knowledge.

### **Syracuse University Certificate of Advanced Study in Data Science**

<https://ischool.syr.edu/future/cas/datascience.aspx>

Fifteen-credit graduate certificate that can be taken as post-baccalaureate. Six credits (two courses) are required in Data Administration Concepts and Database Management and Applied Data Science. Nine credits (three courses) are taken as electives. These courses can be taken both online or in person.

### **University of Illinois at Urbana-Champaign Certificate in Advanced Study in Data Science**

[https://www.lis.illinois.edu/academics/degrees/specializations/data\\_curation](https://www.lis.illinois.edu/academics/degrees/specializations/data_curation)

Five classes must be completed to earn a specialization. Classes are offered through LEEP Online Learning. There are three required courses on metadata, digital preservation, and foundations of data curation. Those currently enrolled in the library school program at Urbana-Champaign may also specialize in data curation.

## **Online Continuing Education and Massive Open Online Courses (MOOCs)**

These online courses provide opportunities for librarians interested in going beyond the library and research data management foci. With the exception of the Library Juice course listed, these courses are intended for those who wish to explore a career in data science. This may be beyond the purview of services that libraries may offer, however, further awareness of the field, the needs of the researchers, and more experience with the software and vocabulary can enhance opportunities in the library.

### **Johns Hopkins University Data Science Specialization**

<https://www.coursera.org/specialization/jhudatascience/1>

This series of nine month-long courses, with a capstone project at the end, is targeted at those interested in learning to be a data scientist. The courses focus on using R statistical software to understand different aspects of data science. Basic programming familiarity is strongly recommended.

### **Library Juice Academy Research Data Management Course**

<http://libraryjuiceacademy.com/082-data-management.php>

The purpose of this asynchronous course is to explore the processes of data production and data management, and the role of LIS professionals and institutions in supporting data producers. Attendees will be able to prepare a data management plan appropriate for submission to NSF/NIH/NEH, a data curation profile, or an institutional data management policy. Course is available for special sessions only. Contact [inquiries@libraryjuiceacademy.com](mailto:inquiries@libraryjuiceacademy.com) to make arrangements. Continuing education units available.

### **Lynda**

<http://www.lynda.com/>

This company provides video courses in software, business skills, and creative skills. These include software classes targeted at specific platforms, such as a Microsoft Office;

programming languages such as Python, or the R statistical language; and more general courses such as an introduction to data visualization. All of the courses are asynchronous and certification of completion is available. Lynda access may be purchased individually or may be available on an institutional level or from the public library.

### **Udacity Data Analyst Nanodegree**

<https://www.udacity.com/course/data-analyst-nanodegree--nd002>

Targeted at an "Intermediate" level learner, this ten month curriculum is targeted at preparing students to be a data analyst. A strong understanding of statistics and programming ability in Python is required before participating.

## **Current Awareness & Networks**

In addition to finding foundational materials and courses to develop research data management skills, it is also important for librarians to stay current with emerging trends in research data management. As there are many options under each of these categories, a general description of the value and benefit is provided with select resources.

### **Journals**

Presently, there is not a journal that solely focuses on research data management and librarianship. While one may emerge in the future, current articles and research can be found scattered across journals from a variety of organizations serving academic, medical, science, and scholarly communications librarians. The majority provide both peer reviewed research as well as topical pieces, special editions, and book reviews relating to research data management. For an extensive recommended reading list, Bailey's *Research Data Curation Bibliography* ([2015](#)) is suggested.

- *College & Research Libraries. (C&RL)* Published monthly by the Association of College & Research Libraries. Open Access.  
<http://crl.acrl.org/site/misc/about.xhtml>
- *College and Research Libraries News. (C&RL News)* Published monthly by the Association of College and Research Libraries. Open Access.  
<http://crln.acrl.org/>

- *D-Lib Magazine*. Published bimonthly by the Corporation for National Research Initiatives. Open Access.  
<http://www.dlib.org/>
- *Data Science Journal*. Published by Ubiquity Press. Published infrequently. Open Access.  
<http://datascience.codata.org/>
- *Evidence Based Library and Information Practice*. (EBLIP) Published quarterly by the University of Alberta Learning Services. Open Access.  
<http://ejournals.library.ualberta.ca/index.php/EBLIP>
- *Information Technology and Libraries*. (ITAL) Published quarterly by the Library & Information Technology Association. Open Access.  
<http://ejournals.bc.edu/ojs/index.php/ital/index>
- *International Journal of Digital Curation*. (IJDC) Published biannually by the University of Edinburgh. Open Access.  
<http://www.ijdc.net/>
- *Issues in Science & Technology Librarianship*. (ISTL) Published quarterly by the Science & Technology Section of the Association of College & Research Libraries. Open Access.  
<http://www.istl.org/>
- *Journal of Academic Librarianship*. (JAL) Published bimonthly by Elsevier. Requires Subscription.  
<http://www.journals.elsevier.com/the-journal-of-academic-librarianship/>
- *Journal of eScience Librarianship*. (JESLIB) Published by the University of Massachusetts Medical School. Open Access.  
<http://escholarship.umassmed.edu/jeslib/>
- *Journal of Librarianship and Scholarly Communication*. (JLSC) Published quarterly by Pacific University Library. Open Access.  
<http://jlsc-pub.org/>
- *Journal of the Medical Librarian Association*. (JMLA) Published quarterly by the Medical Library Association. Open Access.  
<https://mla.sociious.com/p/cm/ld/fid=62>
- *Medical Reference Services Quarterly*. Published quarterly by Taylor and Francis. Requires Subscription.  
<http://www.tandfonline.com/loi/wmrs20#.VZQOnflViko>

- *Practical Academic Librarianship*. Published quarterly by the Special Library Association Academic Division. Open Access.  
<https://journals.tdl.org/pal/index.php/pal/index>
- *Science and Technology Librarianship*. Published quarterly by Taylor and Francis. Requires subscription.  
[http://www.tandfonline.com/toc/wstl20/current#.Vct\\_rbWvHKE](http://www.tandfonline.com/toc/wstl20/current#.Vct_rbWvHKE)

## Data Sources

Sources for data have greatly proliferated. Rather than attempt to be comprehensive, a few major resources and a recommended aggregator are included below for where many data sources can be identified.

### DATA.GOV

<http://www.data.gov/>

Launched in 2009, data.gov was initially intended as a portal to federal government data. It presently includes both state and local data. Over 150,000 datasets covering a broad array of subjects are included for download and reuse.

### Inter-university Consortium for Political and Social Research (ICPSR)

<https://www.icpsr.umich.edu/icpsrweb/landing.jsp>

The ICPSR is a unit of the Institute for Social Research at the University of Michigan. As a consortium of over 700 institutions, it hosts more than 500,000 social science datasets. In addition to hosting and curation services, they offer coursework, educational materials, and workshops to prepare others for doing data review, preparing data for deposit in a trusted repository for reuse, and working with confidential data.

### Re3data Registry of Research Data Repositories

<http://www.re3data.org/>

This registry of research data repositories serves as an excellent gateway for finding data for instruction, education, research and reuse. Over 1,300 repositories are included. Users can browse or search by subject, content type, or by country.

## Library Professional Organizations



Professional organizations offer a variety of ways to stay current with trends in research data management as well as develop peer network who are engaged in similar activities. The organizations offer conferences, in-person and online continuing education, reference material, white papers, and multiple communication venues such as mailing lists. As with the journals, while the organizations may not be fully focused on research data management, the expansion of research data management services has led many of them to create interest groups to further support librarians engaged in these activities.

### **Association of College and Research Libraries. (ACRL)**

<http://www.ala.org/acrl/>

Along with multiple journals already mentioned, there are association-wide interest groups and sections that often have a targeted focus on research data management in practice. The association also offers continuing education such as online classes and preconference workshops at American Library Association conferences. These include:

- Science & Technology Section  
<http://www.ala.org/aboutacrl/directoryofleadership/sections/sts/stswebsite>
- Digital Curation Interest Group  
<http://www.ala.org/acrl/aboutacrl/directoryofleadership/interestgroups/acr-igdc>
- Numeric and Geospatial Data Services in Academic Libraries Interest Group  
<http://www.ala.org/acrl/aboutacrl/directoryofleadership/interestgroups/acr-igngdsal>
- Health Sciences Interest Group  
<http://www.ala.org/acrl/aboutacrl/directoryofleadership/interestgroups/acr-ighs>

### **Association of Research Libraries (ARL)**

<http://www.arl.org/>

ARL provides education, research and support particularly targeted at member libraries on research data management. Many large institutions participated in the ARL/DLF E-Science Institute, which was offered to encourage libraries looking to boost their support of e-research. An especially useful resource is the SPEC Kits, which dive more deeply into a specific area and provide current examples from

libraries. One example is the SPEC Kit 334: Research Data Management Services ([Fearon, et al. 2013](#)) which outlines the activities of current ARL member libraries and provides examples of best practices from those institutions.

### **Medical Library Association. (MLA)**

<https://www.mlanet.org/>

MLA provides much continuing education on research data management, both as workshops at the annual conferences as well as online education. This continuing education is usually certified for professional development credit.

### **Special Libraries Association (SLA)**

<http://www.sla.org>

Because of the diverse nature of SLA, the organization has many geographic chapters and targeted divisions to meet needs of smaller library groups. This includes the Biomedical and Life Sciences, Chemistry, Engineering and other divisions. Divisions and chapters offer targeted continuing education opportunities throughout the year as well as journals and other resource guides.

## **Data Professional Organizations**

In addition to the organizations focused on librarianship specifically, there are a number of information science and data groups that provide continuing education. This broader blend of research and practice and different audience give the opportunity for different perspectives and theories approaching data management.

### **Association for Information Science and Technology (ASIS&T)**

<https://www.asist.org/>

Self-described as a bridge between information science theory and research and practice, ASIS&T is an international group with members that span areas of information and computer science as well as education, linguistics, etc. Of particular interest to librarians is the Research Data Access and Preservation (RDAP) listserv and annual summit. The RDAP summit brings together data librarians for a concentrated discussion across disciplines to mix theory and practice. The listserv is active throughout the year with discussions and ongoing collaboration activities.

## **International Association for Social Science Information Services and Technology (IASSIST)**

<http://www.iassistdata.org/>

Though the original focus of this association was more heavily in social science areas, the organization has expanded focus to more comprehensively include library and information science as well as data science internationally. The primary activity is the annual conference, held in locations around the world, though usually in Europe or North America.

## **Research Data Alliance**

<http://rd-alliance.org/>

This global forum works across international borders to promote open sharing of data around the world, unimpeded by social or technical challenges. There is a formal Libraries for Research Data Interest Group (<https://rd-alliance.org/groups/libraries-research-data.html>). Plenary meetings are held annually.

## **Hands-On Tools**

As librarians are tackling these new services there are software solutions and resource guides available from a variety of sources. These tools will assist librarians in direct researcher support.

## **ACRL Scholarly Communication Toolkit**

<http://acrl.ala.org/scholcomm/>

While not only covering data management topics, this toolkit places data management within the framework of library activities surrounding scholarly communication. Further information is available also about author's rights, scholarly publishing, and digital humanities.

## **Data Carpentry**

<http://www.datacarpentry.org/>

This organization focuses on teaching fundamental concepts surrounding use and management of research data. While the primary focus is on the hosted in-person workshops, they also provide lesson plans released under CC-BY license to guide learners through tools and best practices.

## **DataCite**

<https://www.datacite.org/>

This is an international non-profit organization with a goal of making data citation easier and more widely accepted to improve the status of research data as a scholarly object. The best known product of DataCite is its minting of Digital Object Identifiers (DOI).

## **Data Management Plan Tool (DMP Tool)**

<https://dmptool.org/>

Hosted by the California Digital Library, the DMPTool is designed to walk researchers through the process of creating a data management plan for a grant. Templates and guided questions are available for major funders, including NSF Divisions. Universities can arrange to have their institutional login recognized.

## **e-Science Portal for New England Librarians**

<http://esciencelibrary.umassmed.edu/>

This portal, hosted by the University of Massachusetts, provides bibliographies curated by librarians actively engaged in research data activities. In addition to self-education materials, the portal also includes curated lists of tools, ongoing research, and subject area resources.

## **Purdue Data Curation Profile Toolkit**

<http://datacurationprofiles.org/>

This toolkit was developed to give librarians a question guide to work through with researchers to determine their research data management needs. Profiles can be tailored by discipline and completed examples are available for review.

## **Social Media and Other Self-Education Opportunities**

As social media changes frequently, specific lists of librarians to follow on any certain platform such as Twitter will be nebulous. However, these blogs and other current resources provide ideas about the present state of research data management conversations in social media.

### **Blogs**

**Data Ab Initio**

<http://dataabinitio.com/>

Written by Kristin Briney, PhD, this blog includes current events in research data management and practical suggestions for researchers.

**Databrarians**

<http://databrarians.org>

A multi-author blog targeted at data librarians and MLS students looking to join the field. A number of sources to further engage in the data librarian community are provided.

**Flowing Data**

<http://flowingdata.com/>

Targeted at those looking to better visualize their data, this blog regularly highlights different data visualization techniques, points out interesting examples thereof, and explains common issues.

**KD Nuggets**

<http://www.kdnuggets.com/>

This web site focuses on data mining, analytics, big data, and data science. It provides regular updates about jobs, software, conferences, and other ways to engage with data science.

**Videos****Data Sharing and Management Snafu in 3 Short Acts**

[https://www.youtube.com/watch?v=66oNv\\_DJuPc](https://www.youtube.com/watch?v=66oNv_DJuPc)

Created by the NYU Health Sciences Library, this video quickly demonstrates challenges that may arise when a researcher is asked for their data after the paper has been published.

**University of Wisconsin Data Services Video Series**

<https://www.youtube.com/channel/UC9MT-uDcJJgSXQWC9JvlfeQ>

Created by the UW-Milwaukee Data Services and the UW-Madison Research Data Services group, this series of short video series provides ideas and advice to researchers on topics that are regularly questioned, such as file naming conventions, backing up data, etc. These videos are not tied

to UW resources and can easily be shared with a broader audience.

## **Other Resources**

### **Big Data to Knowledge**

<https://datascience.nih.gov/bd2k>

An overarching NIH initiative that includes software development, training grants, and other research funding. Launched in spring 2015, these grants were given to institutions to develop educational activities and software solutions in Big Data to meet biomedical, behavioral and clinical researchers. The results of these grants will include MOOCs, open education resources, skill development courses, and software solutions that will be available for self-education and reuse.

### **DataQ**

<http://researchdataq.org/>

This new IMLS-funded grant project, launching in Fall 2015, creates a repository of commonly asked questions and provides answers from a panel of data librarian experts.

### **Dorothea Salo's Horror Story Pinboard**

<https://pinboard.in/u:dsalo/t:horrorstories>

Gathered from a variety of web sources, each story highlights an issue of data mismanagement including ownership, sharing, security, policy, publishing, and other challenges.

### **Twitter**

Follow the #datalibs hashtag to discover ongoing conversations and questions

## **References**

**Bailey, C.** 2015. Research Data Curation Bibliography. [accessed 2015 Aug 19]. <http://digital-scholarship.org/rdcb/rdcb.htm>

**Fearon, D.J., Gunia, B., Lake, S., Pralle, B.E. & Sallans, A.L.** 2013. Research Data Management Services, SPEC Kit 334 (July 2013). [accessed 2014 Oct 21]. <http://publications.arl.org/Research-Data-Management-Services-SPEC-Kit-334/>

**National Science Foundation.** 2010. Dissemination and Sharing of Research Results. US NSF - About. [accessed 2015 Jan 4].<http://www.nsf.gov/bfa/dias/policy/dmp.jsp>

**Westra, B.** 2014. Developing Data Management Services for Researchers at the University of Oregon. In: Ray, J.M., editor. Research Data Management: Practical Strategies for Information Professionals. West Lafayette, Indiana: Purdue University Press. p.375-391.

**Whitmire, A., Briney, K., Nurnberger, A., Henderson, M., Atwood, T., Janz, M., Kozlowski, W., Lake, S., Vandegrift, M. & Zilinski, L.** 2015. A table summarizing the Federal public access policies resulting from the US Office of Science and Technology Policy memorandum of February 2013. [accessed 2015 May 18].[http://figshare.com/articles/A\\_table\\_summarizing\\_the\\_Federal\\_public\\_access\\_policies\\_resulting\\_from\\_the\\_US\\_Office\\_of\\_Science\\_and\\_Technology\\_Policy\\_memoandum\\_of\\_February\\_2013/1372041](http://figshare.com/articles/A_table_summarizing_the_Federal_public_access_policies_resulting_from_the_US_Office_of_Science_and_Technology_Policy_memoandum_of_February_2013/1372041)

[Previous](#)

[Contents](#)

[Next](#)



This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

