

2016

Introduction to Databrarianship: The Academic Data Librarian in Theory and Practice

Kristi Anne Thompson
University of Windsor

Lynda Kellam
University of North Carolina at Greensboro

Follow this and additional works at: <http://scholar.uwindsor.ca/leddylibrarypub>



Part of the [Scholarly Communication Commons](#), and the [Social Statistics Commons](#)

Recommended Citation

Thompson, Kristi Anne and Kellam, Lynda. (2016). Introduction to Databrarianship: The Academic Data Librarian in Theory and Practice. *Databrarianship: The Academic Data Librarian in Theory and Practice*.
<http://scholar.uwindsor.ca/leddylibrarypub/47>

This Contribution to Book is brought to you for free and open access by the Leddy Library at Scholarship at UWindsor. It has been accepted for inclusion in Leddy Library Publications by an authorized administrator of Scholarship at UWindsor. For more information, please contact scholarship@uwindsor.ca.

Introduction

Kristi Thompson and Lynda Kellam

Editors

With the appearance of big data, open data, and particularly research data curation on many libraries' radar screens, data services has become a topic of increasing interest in academia, but in academic libraries, this area has long been a central concern. The term "databrarian" dates to a 2013 *Library Journal* article that discussed the results of the 2012 Placements & Salaries survey and noted "several new job titles," including Research Data Librarian, Data Coordinator, and Data Curation Specialist.¹ While practitioners welcomed the increased recognition, data librarianship as a field is hardly new, as it encompasses a diversity of forms, functions, and specializations that are vital to academic research and teaching. Drawing on the expertise of a diverse community of practitioners, the goal of this edited volume is to provide an overview of the major areas within data librarianship. By covering the data lifecycle from collection development to preservation, examining the challenges of working with different forms of data, and exploring service models suited to a variety of library types, this volume provides a toolbox of strategies that will allow librarians and administrators to respond creatively and effectively to the data deluge.

In this collection, we use the term databrarianship to describe a field that is characterized by a diversity of interests but united by our commitment to ensuring access to data, be they primary research data created by our institutions' researchers or secondary data used for analysis. Our contributors have widely diverse primary positions and backgrounds. Some are reference librarians who support data access, which requires a fluency with data concepts and sources not typically taught in the library school reference class. Some are data curators who share more in common with metadata specialists and bring a unique understanding of preservation methods for

long-term data access. Some are library instructors who teach data source discovery and use datasets and statistics as part of their library's larger information literacy program. Several are data support specialists without MLIS degrees who work in research grant or IT offices and provide unique insights into the support of researchers. All have a passion for data and for supporting our faculty and students.

Although this specialty area has been around for many years, relatively few texts have been written exclusively about it, and in many cases these publications have remained within the data librarian community. During the summer institute at the Inter-university Consortium for Political and Social Research (ICPSR), a leading data archive, many budding data librarians were introduced to the field by Diane Geraci, Chuck Humphrey and Jim Jacobs and their excellent textbook, *Data Basics: An Introductory Text*.² In addition, *IASSIST Quarterly*, the journal of the International Association for Social Science Information Services and Technology (IASSIST), the leading professional organization for data services, has been a key resource for social science data support since 1976. In both cases, these publications have primarily reached audiences already involved in data librarianship and support. Beyond these key texts, little has been written for a wider audience. A few major figures such as Geraci, Jacobs, Judith Rowe, and others wrote articles in the 1980s and 1990s discussing the possibility of bringing data files (at that time on magnetic tape) into library collections and data support into the skill set of reference librarians.³ While these articles were featured in general librarianship journals, they were few and far between. In the mid-2000s, increased interest in data and access to data through the internet led to a growth in the number of articles and books about supporting data services. Eleanor Read's 2007 article in *Reference and User Services Quarterly* helped many new data services librarians have a better understanding of support for data and assessment of services.⁴

Moreover, the 2011 volume *Numeric Data Services and Sources for the General Reference Librarian* by Lynda Kellam and Katharin Peter aimed to introduce a wider audience to support for data discovery and instruction.⁵ In recent years, more volumes have emerged focusing on research data management with Jake Carlson (formerly of Purdue University) taking a lead with the Data Information Literacy Toolkit, a toolkit for having discussions about data with faculty, and the subsequent book *Data Information Literacy: Librarians, Data, and the Education of a New Generation of Researchers* in 2014.⁶ While all of these works have filled necessary niches, no one work has tried to span the breadth of the primary concerns of the profession for a general audience. With this collected edition, our authors explore more traditional areas like data reference as well as the emerging area of data management. We hope this text provides an introduction to the scope of databrarianship as well as the variety within the profession.*

Data librarian and archivist positions emerged in response to the varied needs of researchers and data producers, and at first these professionals organized and communicated informally through a variety of venues. A major event in the history of databrarianship was the Conference on Data and Program Library Services that was held in conjunction with the World Sociology Congress meeting in Toronto in 1974. This meeting led to the founding of the IASSIST by librarians and archivists from Canada, the U.S., and Europe.⁷ The influence of IASSIST can be seen to this day and is evident in this book in a two major ways. First, Americans, Canadians, and Europeans played key roles in the early development of the profession and its identity, in collaboration and dialogue with each other. Consequently, the editors made an effort to include voices from beyond the American experience. A second factor is the roots of databrarianship as a profession in the quantitative social sciences. The astute

* Access links, extra appendices, and more on our companion website: <https://databrarianship.wordpress.com/>

reader will note that a number of chapters, particularly in the section on services, focus on quantitative survey data. To ensure that other forms of data were covered, we recruited authors to write about data from other disciplinary focuses.

Despite this effort to include a variety of perspectives, much is still missing. One clear gap is that we were unable to cover the full range of international experiences. Descriptions of data support in countries outside of the U.S., Canada, and Western Europe are missing. This does not mean that data librarians are not active in these countries, and we are aware that excellent work is being done by data librarians in places as diverse as South Africa, Japan, and Croatia, and undoubtedly many others. The work experiences of these data librarians may be quite different from the position of the data librarian in North America, as Rice's chapter discusses.[†] We hope that other publications will investigate these differences and develop a more global perspective on databrarianship.

Before moving into our chapters, however, this introduction must lay the groundwork for what is to come. Those new to data librarianship will encounter many new terms throughout this book, and our authors do an excellent job of providing concise definitions. One critical term must be discussed first though, and that is the word *data*. The term *data* is used in a variety of ways depending on field and context. A computer scientist might use the term to refer to the flow of zeros and ones that stream through data cables and are used to transmit videos, web pages, and the like, or to the aggregate of all the information available on the internet. A statistician or survey researcher might think of a numeric dataset structured for use in a statistical package, such as the results in a public opinion poll. In general, we take an intermediate approach. The data we are concerned with here are the product of taking that raw informational input and

[†] For more information about data support in Great Britain and Europe, see the chapter in this volume by Robin Rice called "View from Across the Pond."

assembling it into a structured form for analysis. Data are a product of research as well as an input for research. Research data collections (or *datasets*) are generally in electronic form and are accompanied by or incorporate metadata, or documentation that describes the structure and content of the data. In brief, unless specified otherwise, data will be taken to mean electronic files containing information that has been collected systematically, structured, and documented to serve as input for further research. Data are the raw materials for research, produced through any systematic collection of information for the purpose of analysis. Data are the lifeblood of data librarianship, but as you can see in this book, the term does not belong to one discipline or one research methodology, and we will cover the wide range of interests and concerns in our profession.

We have grouped the chapters into four major sections that give some idea of the division of interests within the field: *Data Support Services for Researchers and Learners*, *Data in the Disciplines*, *Data Preservation and Access*, and *Data: Past, Present, and Future*. The first section, *Data Support Services for Researchers and Learners*, is intended to serve as a toolbox for developing academic library services for data creators and consumers, encompassing data-focused perspectives on familiar librarian activities, such as reference and information literacy, and also branching out to cover topics specific to databrarianship, such as data wrangling and support of researchers working with restricted data. We begin with a pair of chapters that contrast the merits of two different modes of service provision. Samantha Guss describes offering a suite of functions at a tightly focused service point in “The Data Studio Model,” while Cynthia Hudson considers how one might instead integrate as a databrarian seamlessly into the campus at point of need with “Embedded Options.” Our next pair of chapters offers another study in contrasts, this time comparing top-down and bottom-up techniques for service

development. Bobray Bordelon's "Data Reference" offers advice and encouragement for individual subject specialists looking to incorporate reference for microdata into their job descriptions, with or without the assistance of a dedicated data librarian, while Alicia Hofelich Mohr, Lisa Johnston, and Thomas Lindsay take a systemic approach to configuring campus-wide services in the "Data Management Village." Next, Ryan Clement looks at options for libraries working on a much smaller scale, demonstrating that specialized data services incorporating reference, management, and statistical consulting are not out of reach of "The Databrarian in the Liberal Arts College." The role of librarians in teaching novices is not neglected, as Adam Beauchamp and Christine Murray consider the needs of beginning data users in their information literacy-focused take on "Teaching Foundational Data Skills in the Library," which is followed by a discussion of technical skills with Harrison Decker and Paula Lackie's "An Introduction to Data Wrangling for Reproducible Research." Finally, Jen Darragh takes on the needs of researchers using restricted access data, a type of data that requires protections such as a secure facility or computer. As Darragh demonstrates in "Restricted Data Access and Libraries," the library may be uniquely positioned to help researchers use this type of data.

Data in the Disciplines ventures across the campus to look at a few of the many specialty areas of concern for data support. Our first two chapters tackle geospatial data, a popular and, for some, newer area of support. Nicole Scholtz begins a discussion of "Supporting Geospatial Data" by looking at the wider context of data services and provides suggestions for libraries with an interest in developing their geospatial services. Joy Suh follows with an in-depth case study of geospatial academic support that looks at how a service evolves to meet advanced user needs in "From Traditional to Crowd and Cloud: Geospatial Data Services at GMU." Next, Mandy Swygart-Hobaugh gives qualitative research much-needed attention—and encourages other

librarians to do the same—in her survey chapter “Qualitative Research and Data Support: The Jan Brady of Social Sciences Data Services?” Karen Grigg rounds out the section with an overview of the unique needs and challenges of providing support for “Data in the Sciences.”

The third section, *Data Preservation and Access*, looks at the place of data in scholarly communication and delves further into the growing field of research data management. Hailey Mooney kicks things off with an exploration of the integral role of data in the production of scientific knowledge in “Scholarly Communication and Data.” Rob O’Reilly and Joel Herndon next present the results of a comparative study of the “Data Sharing Policies in Academic Journals,” focusing on how the policies of leading social science journals have evolved over time. Collection development policies and procedures are the focus of Christopher Eaker’s “Appraisal and Selection of Digital Research Data,” and Susan McKee has a collection development case study in “Local Data Success Story: The University of Calgary Library’s Ten Years with the City of Calgary.” The concluding two chapters in this section consider the different forms that metadata can take and how it can best be used for data preservation and dissemination. Amber Leahey and Jane Fry consider how data professionals can collaborate to produce structured metadata for social science data, specifically the popular standard known as DDI in “Metadata for Social Science Data and Collaborative Best Practices.” Finally, Lizzy Rolando, Lisha Li, Ameet Doshi, Alison Valk, and Karen Young take an opposite approach and investigate ways of encouraging individual users to develop unstructured, discipline-specific metadata in “Exploring Disciplinary Metadata and Documentation Practices to Facilitate Data Reuse.”

We conclude with *Data: Past, Present, and Future*. Two chapters examine the history of how data librarianship developed in two very different sets of circumstances. Robin Rice tells the

story of the European context in “View from Across the Pond: A UK Perspective on Data Librarianship,” while Vincent Gray and Elizabeth Hill return us to North America in “Canadian Data Librarianship: A Historical Perspective.” Dorianne Mizzy examines the many career paths librarians and informationists take when developing as data services professionals in “Data Librarianship: A Day in the Life—Science Edition.” Concluding the volume, Walter Giesbrecht and Michael McCaffrey look to the future as they describe a course designed to teach the fundamentals of databrarianship to a new crop of library students and lay out a proposal for a full curriculum in “Teaching Data Librarianship to LIS Students.”

By now it should be clear that academic data librarianship is not a single specialty but rather a varied collection of overlapping but distinct roles that center on providing access to, documenting, and preserving data, much as traditional librarianship has done for print resources. Most databrarians will share certain base skill sets, having knowledge of file formats, documentation and metadata standards, and disciplinary research practices, but beyond that, their jobs can take a myriad of forms. Underlying many of our chapters is the idea that data librarianship is a collaborative endeavor across our libraries and our universities. The needs of our patrons are diverse as are the skill sets required to assist them. As editors, our motivation with this collection is to demonstrate the wide breadth of data librarianship and all of the component positions. From data discovery to data instruction techniques to data curation and metadata knowledge, databrarianship is a broad field with a world of opportunity for collaboration and innovation. We hope this volume provides some insight into that world.

¹ Stephanie L. Maatta, “The Emerging Databrarian,” *Library Journal* 138, no. 17 (Oct 15, 2013): 26.

² Diane Geraci, Chuck Humphrey, and Jim Jacobs, *Data Basics: An Introductory Text*, (Ann Arbor, Michigan: ICPSR, 2008).

³ Judith Rowe, “Expanding Social Science Service to Meet the Needs of Patrons More Adequately,” *Library Trends* 30, no. 3 (Winter 1982): 327-334; Kathleen M. Heim, “Introduction: Data Libraries for the Social Sciences,” *Library Trends* 30, no. 3 (Winter 1982): 321-325; Ray Jones and Colleen Seale, “Expanding Networks: Reference Services for MRDF,” *Reference Services Review* 16, no. 1 (July 8, 1988): 7-12; Deborah Rinderknecht, “Non--

Bibliographic Databases: Determining Level of Service,” *RQ* 30, no. 4 (July 1991): 528-533; James A. Jacobs, “Providing Data Services for Machine-Readable Information in an Academic Library: Some Levels of Service,” *The Public-Access Computer Systems Review*, 2 (1992): 119-132; Diane Geraci and Linda Langschieid, “Mainstreaming Data: Challenges to Libraries,” *Information Technology and Libraries* 11, no. 1 (March 1992): 10-19; Wendy Treadwell and James A. Cogswell, “The Machine Readable Data Center: A Model Approach to Data Services in Academic Research Libraries,” *Library Hi Tech* 12, no. 1 (1994): 87-92.

⁴ Eleanor J. Read, “Data Services in Academic Libraries: Assessing Needs and Promoting Services,” *Reference & User Services Quarterly* 46, no. 3 (2007): 61-75.

⁵ Lynda M. Kellam and Katharin Peter, *Numeric Data Services and Sources for the General Reference Librarian*, (Oxford: Chandos Publishing, 2011).

⁶ Jake Carlson and Lisa R. Johnston, *Data Information Literacy: Librarians, Data, and the Education of a New Generation of Researchers*, (West Lafayette, IN: Purdue University Press).

⁷ For more on the early history of data librarianship, see Margaret O’Neill Adams, “The Origins and Early Years of IASSIST,” *IASSIST Quarterly* 30 (Fall): 5-14, <http://www.iassistdata.org/sites/default/files/iq/iqvol303adams.pdf>.