

2017-10

Evolution of the Data Librarian

Tschirhart, Lori

<https://hdl.handle.net/2027.42/139704>

<http://creativecommons.org/publicdomain/zero/1.0/>

Downloaded from Deep Blue, University of Michigan's institutional repository



GEOSCIENCE INFORMATION SOCIETY

Four Years of Earth Science Information: Exploring Data, Access, and More



**Proceedings of the 48th, 49th, 50th, and 51st Meetings
of the Geoscience Information Society**

**October 26–30, 2013
Denver, Colorado, USA**

**October 18–22, 2014
Vancouver, British Columbia, Canada**

**October 31–November 3, 2015
Baltimore, Maryland, USA**

**September 24–28, 2016
Denver, Colorado, USA**

**Four Years of Earth Science Information:
Exploring Data, Access, and More**

**Edited by
Matt Hudson**

Proceedings

Volume 44

© 2017

Geoscience Information Society

www.geoinfo.org

Copyright 2017 by the Geoscience Information Society

Material published in this volume may be reproduced or distributed in any format via any means by individuals for research, classroom, or reserve use.



GEOSCIENCE
INFORMATION
SOCIETY

ISBN: 978-0-934485-45-6

ISSN: 0072-1409

For information about copies of this proceedings volume or earlier issues, visit www.geoinfo.org or contact:

President
Geoscience Information Society
C/O American Geosciences Institute
4220 King Street
Alexandria VA 22302-1502 USA
president@geoinfo.org

Cover photos (from left to right): The big blue bear outside the Colorado Convention Center in Denver, Colorado; west entrance lobby of the Vancouver Convention Center, by Sara Borck Photography, courtesy of Vancouver Convention Center; The George Peabody Library in Baltimore, Maryland, photo courtesy of Shaun Hardy; the exhibit floor at the Colorado Convention Center, Denver, Colorado, photo courtesy Geological Society of America.

TABLE OF CONTENTS

| | |
|--|------------|
| PREFACE | <i>xiv</i> |
| PART 1: 2013 GSA POSTER SESSION T144. Global Vision: Geoscience Information for the Future | 1 |
| GETTING DOWN AND DIRTY: CITIZEN INVOLVEMENT IN SCIENCE Cynthia L. Prosser and Monica Pereira | 2 |
| SHARE AND SHARE ALIKE: USING WIKIMEDIA COMMONS TO DISSEMINATE GEOPHOTOGRAPHY Michael C. Rygel | 7 |
| THE NATIONAL GEOLOGIC MAP DATABASE—A RESOURCE FOR GEOLOGIC MAPPING David R. Soller and Nancy R. Stamm | 8 |
| SCIENTISTS AS MEDIA RESOURCES IN THE AFTERMATH OF DISASTERS: TRENDS IN NEWS COVERAGE FOLLOWING TWO DEVASTATING TSUNAMIS Ellen K. Buelow, Xai Her, and Scott K. Clark | 9 |
| THE PAST IS THE KEY TO THE FUTURE: URANIUM RESEARCH AT THE USGS DENVER LIBRARY Emily C. Wild and Keith Van Cleave | 10 |
| BACK TO THE FUTURE: URANIUM INFORMATION AT THE USGS DENVER LIBRARY Emily C. Wild and Keith Van Cleave | 11 |
| GEOLOGICAL, ATMOSPHERIC, CLIMATIC DATA: COLLECTIONS IN THE U.S., POSSIBLY IN THE WORLD. A VISUAL PRESENTATION OF TYPES OF DATA, FORMATS, AND LOCATIONS OF STORAGE Louise Deis | 12 |
| DATA CITATION AND METRICS IN THE GEOSCIENCES Amanda Bielskas | 13 |
| BRINGING HISTORIC MAPS TO LIFE: GEOREFERENCING FOR THE DIGITAL GLOBE TO SUPPORT INTERDISCIPLINARY SCHOLARSHIP Katherine H. Weimer and Bruce E. Herbert | 19 |
| PART 2: 2013 GEOSCIENCE INFORMATION SOCIETY EVENTS | 20 |
| SCHEDULE OF EVENTS | 21 |
| “GEOSCIENCE LIBRARIANSHIP 101” | 22 |
| 2013 BUSINESS MEETING MINUTES | 23 |

| | |
|---|------------|
| DECISION MATRIX FOR DATA SHARING AND STORAGE Linda R. Musser | 107 |
| A WEB-BASED TOOL FOR PREPARING FGDC-CSDGM METADATA Ranjeet Devarakonda and Giri Palanisamy | 110 |
| SIMILARITIES IN DIFFERENT NATURES—A COMPARISON OF SOCIAL MEDIA USE BETWEEN HURRICANE SANDY AND OSO MUDSLIDE Xiangyang Guan and Cynthia Chen | 111 |
| UTILIZING DIGITAL MEDIA TO COMMUNICATE SCIENTIFIC FINDINGS WITH A DIVERSE AUDIENCE Cristina M. Robins, Claudia Grant, Aaron R. Wood, Shari Ellis, Bruce J. MacFadden | 112 |
| PART 6: 2014 GEOSCIENCE INFORMATION SOCIETY EVENTS | 113 |
| SCHEDULE OF EVENTS | 114 |
| “GEOSCIENCE LIBRARIANSHIP 101” | 115 |
| 2014 BUSINESS MEETING MINUTES | 116 |
| SUMMARY OF 2014 PROFESSIONAL ISSUES ROUNDTABLE Michael Noga | 126 |
| GSIS MEMBER & NON-MEMBER (PRELIMINARY) STRATEGIC PLANNING SURVEY RESULTS Amanda Bielskas | 128 |
| 2014 AWARD WINNERS | 139 |
| PART 7: 2015 GSA TECHNICAL SESSION T95. Evolution of the Data Life Cycle | 141 |
| OVERVIEW OF EMERGING REQUIREMENTS FOR DATA MANAGEMENT OF FEDERALLY FUNDED RESEARCH IN THE GEOSCIENCES Richard Huffine | 142 |
| EVOLUTION OF THE DATA LIBRARIAN: DOCUMENTING THE EVOLVING ROLE OF DATA LIBRARIANS AT THE UNIVERSITY OF MICHIGAN LIBRARY Lori Tschirhart | 143 |
| THE EVOLUTION, APPROVAL, AND IMPLEMENTATION OF THE U.S. GEOLOGICAL SURVEY SCIENCE DATA LIFECYCLE MODEL John Faundeen, Tom Burley, David Govoni, Heather S. Henkel, Elizabeth Martin Ellyn Montgomery, Lisa Zolly, Viv Hutchison, Steve Tessler, and Cassandra Ladino | 146 |
| WHO IS USING OUR DATASETS AND DATA PRODUCTS? Carol A. Deering | 147 |

EVOLUTION OF THE DATA LIBRARIAN: DOCUMENTING THE EVOLVING ROLE OF DATA LIBRARIANS AT THE UNIVERSITY OF MICHIGAN LIBRARY

Lori Tschirhart

University of Michigan Library

919 South University Ave., Ann Arbor, MI 48109-1185, ltz@umich.edu

Abstract—Data services provided by libraries have grown and changed over time. This paper will explore the evolving services, priorities, and capacities of data librarians at one ARL library in relation to evolving demands by university stakeholders. The paper will also consider the many skill sets required in recent data librarian position descriptions, and will seek to clarify the most crucial skills necessary to fulfilling the research data needs of the earth sciences community.

Introduction

University of Michigan Librarians have provided services related to Geographic Information System (GIS) in some fashion or other for over twenty years. Early on, a group of librarians developed relationships with the geographer community and social science researchers due to their work within the library's robust maps collection and ties to the local Inter-University Consortium for Political and Social Research (ICPSR). More recently, demand for GIS support has become more prevalent among diverse disciplinary communities. To address increasing demand, new roles have been formally developed over time for Spatial and Numeric Data (SAND) librarians. These SAND librarians are often referred to as "functional experts" for the deep and specific expertise they offer related to spatial and numeric data. Because they have been a distinct service point for select data services through many years, I refer to them as "old school" data librarians.

Very recently, traditional subject librarians at the same institution have modified their research consultation services on a system-wide level to offer new modes of support for scholars using or creating research data. These new services are intended to support research at every phase of the research data lifecycle, from data management planning to curation and preservation.

In an effort to develop new research data services complementary to established spatial and numeric data services, and to better understand the specialized expertise held by SAND librarians, I conducted hour-long interviews with multiple librarians con-

nected to spatial and numeric data service offerings. I asked questions designed to help me elicit details regarding their distinct areas of expertise and how they were first articulated and championed. I also invited these colleagues to discuss potential areas for growth and barriers to maximizing the value of their expertise related to spatial and numeric data. Through this paper, I will describe their responses, from their understanding of the early impetus for traditional data librarian services to their speculation on the future of our library's old school data librarian services.

Historical Data Librarian Services

In the early 1990s, the geographer community served by the library was increasingly interested in using GIS tools thanks to development and dissemination of digital spatial data. GIS and spatial data initiatives were developing, and the library, as a recognized data provider, collaborated. Librarians started supporting ESRI/Census classes, provided access to GIS software and associated training, and facilitated access to data for users. As time went on, more demand for data services grew. GIS assistance was articulated by scholars interested in population studies, health sciences, climatology/meteorology, and remote sensing of natural sciences data. The library responded to growing demand by developing new data librarian positions. The specialized expertise necessary to meet emerging demands began with a half-time librarian position. Over time, as the utility for spatial data became more apparent to the scholarly community, more SAND capacity was added through the development of new librarian positions.

Now, in 2015, the service has grown to employ two full time SAND librarians along with one visualization librarian. Librarians also developed two SAND labs in close proximity to map and architecture collections to provide specialized support to communities engaged with early use of GIS tools.

The development of specific SAND services focused on user needs not met by existing campus services. For instance, statistical analysis services were already available through the University of Michigan's Consulting for Statistics, Computing, and Analytics Research (CSCAR) and so were not replicated within the library. Early on, SAND librarians provided critical assistance with statistical code. Activities included assistance with finding variables, finding and re-purposing code, and reading manuals when users were stuck. Demand for spatial and census data was high, and services skewing toward social sciences researchers were emphasized.

The type of assistance offered by U-M SAND librarians has always involved some negotiation. Even today activities are fluid, varied, and often iterative. The intention and capacity of SAND services do not allow a production shop environment. SAND librarians regularly help people: articulate their data need, acquire data, construct datasets, clean or reduce noise so that data can be used for a distinct purpose, apply methodology, and interpret data. They also offer help with statistics syntax, statistical metrics, data formats, and map making when needed. They provide expertise with common tools including STATA, SPSS, SASS, and R, and provide assistance with tool selection and training. Finally, SAND librarians have long taken responsibility for maintaining knowledge of the location of publicly available data.

In addition to the services described previously, SAND librarians are contributing to the development of an Open Access portal which provides metadata and points to publicly available geospatial resources. At a later phase the portal may also provide metadata and access to licensed archive data at the library.

Emerging Opportunities for Education and Service Development

Over the last ten years, the landscape for spatial and

numeric data services has changed dramatically. Geospatial computing possibilities have changed thanks to technological advances (especially mobile computing). More data is available and more users know how to find it. Campus data users know how to use software more and more often. Research is beginning to scale larger with high performance computing and more complex uses of data are emerging. And new platforms for instructional delivery offer ways to think about better education for future generations of SAND service users.

New developments in geospatial computing consumer services, while welcome, can unintentionally lead inexperienced researchers to believe that they can use the tools and data sets easily and without training. The library can have a role in deliberately defining services and providing education to counter this assumption. Services and instruction can be designed to provide people with chances to realize that the data they have selected doesn't always tell them what they think it is telling them. Our functional specialists have the expertise to highlight differences between using geospatial data and using it well.

The library also has an opportunity to provide new educational content to help scholars develop laboratory skills such as use of version control systems essential for research computing. Many of these skills can be delivered through software carpentry and data carpentry workshops. Software carpentry workshops introduce attendees to software used in laboratory research. Data carpentry workshops provide researchers with domain-specific data skills relevant to analysis and management of research data. This type of expertise is in high demand throughout campus, with many graduate students on research teams specifically sought by departmental research teams for their ability to teach about tools and methods used in their labs. With increased staffing, the library could develop dedicated staff positions to participate with campus partners in the delivery of training for software and data carpentry workshops.

Data visualization can help all kinds of scholars communicate their findings and advance understanding, and can be especially useful in making sense of big data. Data visualization services are already offered at the U-M Library and a new visualization

librarian position has been formalized. Growing demand is anticipated.

SAND librarians have identified a need for education and consultation around the ethics of data re-use. These experts are well poised to raise awareness of spurious correlations and the inadequacies of many statistical models for emerging research. There may be opportunity to mitigate the reproducibility problems of research through discussion, education, and advocacy for new models of publication and scientific experimentation. There is also ripe opportunity to enhance information literacy for future generations by highlighting examples of research and providing ethical guidance.

There has long been a desire among campus librarians for a campus data repository and library-supported data curation. At long last, the library is close to rolling out such a service. Depending on the eventual functionality of this emergent repository, new datasets produced by campus scholars will be described and made available for re-use. Data librarians and their disciplinary expert colleagues will have new opportunities to assist one another in order to support more collaborative research trending on campus.

Perennial Needs

In order to provide services that remain relevant to users, continuous investment in staff training and resources is essential. Librarians need to purchase,

negotiate access to, point to, and host data. Time and skill is also required to manage servers and generally figure out the best ways to provide data and data services. Ongoing investment in infrastructure, too, is necessary to start providing appropriate access to sensitive data and to enable appropriate stewardship of research data. Our library has historically avoided dealing with sensitive data due to lack of infrastructure. All expertise benefits from regular upkeep. Since no expert can be equally good at all things, data librarians need capacity for building expertise. Thoughtful job posting design and avoidance of “kitchen sink” position descriptions will enable functional experts to cultivate deep skill and value for the research community served.

Finally, a down-to-earth mindset about the limitations of data can also bolster the relevance of data librarian services. There can exist a disconnect between the talents of data librarians drawn to complex inquiry and the wishes of some data consumers for quick, clean answers. For this reason, the library must identify and challenge unrealistic demands on the utility of research data and research data services.

Acknowledgments

Many thanks are offered to the following librarians who shared their time, ideas, and institutional knowledge with me and allowed me to synthesize it all into this report: Catherine Morse, Justin Joque, Mara Blake, Nicole Scholtz, Karl Longstreth, Jennifer Green, and Jake Carlson.