The Journal of Academic Librarianship 40 (2014) 541–549

Contents lists available at ScienceDirect

The Journal of Academic Librarianship

REVIEW

# Data Sharing and Discovery: What Librarians Need to Know

Don MacMillan *

*University of Calgary, Libraries & Cultural Resources TFDL 460D, 2500 University Drive NW, Calgary Alberta T2N 1N4, Canada*

ARTICLE INFO

*Article history:*
Received 17 March 2014
Accepted 26 June 2014
Available online 18 July 2014

*Keywords:*
Data sharing
Data discovery
Data repositories
Data literacy
Academic librarians

ABSTRACT

Drawing from the literatures of science, scholarly communication, and librarianship, this review paper describes what librarians need to know about how scientists manage and share their data. It is intended to help librarians become more engaged and integral partners in research and education. Scientific data repositories, journal data deposition policies, and the development of persistent linking between scholarly publications and data sets, have made data more accessible. However, deposition and sharing practices still vary among researchers, journal publishers, data repositories, information providers, and universities. Understanding the dynamic relationships between these stakeholders is critical to providing relevant support to researchers and students in the sciences. Librarians need to develop skills that bridge traditional liaison work with the increasingly data-driven demands of scientific research, so that we can support researchers with their data management needs and help users discover data across myriad collections and resources.

© 2014 Elsevier Inc. All rights reserved.

**CONTENTS**



## INTRODUCTION

Data and data sets have always been integral to scientific research, but in the past two decades various factors have brought them out of the lab notebooks and into publicly available spaces. In many scientific disciplines research has become much more data-intensive and collaborative as a result of innovations in the production and storage of large data sets (Hey, Tansley, & Tolle, 2009; Science Staff, 2011). Disciplines that generate large volumes of data include astronomy (Sloan Digital Sky Survey), physics (Large Hadron Collider), life sciences (Human Genome Project), and climate science (NOAA's Climate Data Center). Data sets underpinning scholarly publications were once just the mostly invisible background for scholarly endeavor that focussed on articles and patents for dissemination. Increasingly they are recognized and disseminated as scholarly output in their own right, and in some disciplines, may be more important to the research agenda than the associated published literature (Akers & Doty, 2013; Castelli,

* Tel.: +1 403 210 8632.
*E-mail address:* macmilld@ucalgary.ca.

http://dx.doi.org/10.1016/j.acalib.2014.06.011
0099-1333/© 2014 Elsevier Inc. All rights reserved.

Manghi, & Thanos, 2013; Reilly, Schallier, Schrimpf, Smit, & Wilkinson, 2011). As Alan Blatecky, the director for Advanced Infrastructure (ACI) at the National Science Foundation, states: "data is the new currency for research" (Markoff, 2013 para. 10).

Sharing data allows other researchers to examine results and reproduce experiments, essential activities in scientific research (Hanson, Sugden, & Alberts, 2011; Roche et al., 2014; Thessen & Patterson, 2011; Vision, 2010). Data sharing in the sciences has received considerably more attention than in the arts or humanities disciplines, due to the data-intensive nature of its research activities and the longer history of formal data sharing mandates from funders, institutions, open access publishers, and fellow researchers (Akers & Doty, 2013; Kim, 2013). While the scientific community endeavors to make data accessible in standardized formats, researchers face many challenges coping with the quantity and complexity of data that is being created. Researchers recognize the benefits of sharing their data, but many are reluctant to do so due to disciplinary or cultural practices, insufficient preservation infrastructure, lack of academic credit or reward incentives, and the perceived effort or cost required to do so (Borgman, 2012; Hanson et al., 2011; Reilly et al., 2011; Tenopir et al., 2011). Other barriers include a lack of domain-specific data repositories or common data citation standards in some disciplines, and ethical or confidentiality considerations in others. Further complication results from collaborative work where some researchers may not have the same disposition toward sharing, or the same institutional or national mandates around sharing and confidentiality (Akers & Doty, 2013; Mooney & Newton, 2012; Roche et al., 2014; Thessen & Patterson, 2011). In spite of these challenges, several initiatives and solutions have emerged that promise to facilitate the discovery and reuse of research data.

The need for more support to address these challenges is critical. Most prominently, the desire for greater accountability has resulted in national, organizational, and institutional mandates for data sharing. Other factors include the increasingly collaborative and interdisciplinary nature of research, and the desire to use resources, including already collected data, more efficiently. Researchers have also realized that data stored in older formats are no longer accessible and that some large collections of data may be vulnerable to loss of funding. While recent initiatives such as those described in this paper may lead to a more stable data environment, science librarians need to be aware of what data are available, where the data are housed and how their users can access the data, much in the same way they have developed knowledge to help users find bibliographic resources most efficiently.

This paper brings together information from sources in the sciences, the publishing industry, and the library field to provide a cohesive overview of data-sharing. It provides context by setting out the fundamental need for data sharing in science, some critical factors both technical and cultural, that inhibit this sharing, and the benefits of overcoming those challenges. It then examines the roles of stakeholders including journal publishers, data repositories, information providers such PubMed, and librarians in making scientific data more accessible to other users. Examples of data sharing initiatives from scientific journal publishers and data repositories are provided to illustrate how some information providers are facilitating data sharing. Throughout the article, the focus is the impact data sharing and related initiatives have on academic libraries, and the conclusion will highlight the types of support academic librarians can provide to our campus partners.

## DATA SHARING

Data sharing is interpreted and practiced differently across disciplines, but Borgman has defined it most simply as "the release of research data for use by others" (Borgman, 2012, p. 1060). Another study describes data sharing as encompassing activities such as attaching data sets to scholarly articles, depositing data sets in repositories, or saving data on a personal computer or local server (Wallis, Rolando, & Borgman, 2013). The effectiveness and value of data sharing vary by discipline or sub-discipline as do the types of data, their life cycle, and the scholarly output which are produced at different rates across disciplines (Wallis et al., 2013).

Hanson states "It is obvious that making data widely available is an essential element of scientific research" (Hanson et al., 2011, p. 649). Data sharing enables researchers to reproduce and validate research results, disclose these results, examine new hypotheses, identify any methodological errors, minimize duplication of resources, and ensure the sustainability and integrity of stored data (Borgman, 2012; Hanson et al., 2011; Reilly et al., 2011; Tenopir et al., 2011). While the importance of making data available to others has always been well understood, distributed databases, robust networks, and relatively inexpensive computer processing power and memory have made it possible to more fully realize the ideals of the past. As Angel Gurria notes in the introduction to the OECD (2007) report on data sharing:

> "Besides, access to research data increases the returns from public investment in this area; reinforces open scientific inquiry; encourages diversity of studies and opinion; promotes new areas of work and enables the exploration of topics not envisioned by the initial investigators".
>
> [Gurria, 2007, p. 3]

These intrinsic motivations for sharing data have been reinforced by requirements from other stakeholders. Increasingly, researchers are required to engage in effective data sharing practices as a result of policies from major government funding agencies such as those in Europe, the US, Australia and Canada. As well, prominent scientific journal publishers have revised their formal data sharing policies to ensure authors deposit supplemental data sets in appropriate domain or community related data repositories, usually as a condition of publication.

Despite these intrinsic and extrinsic motivators for sharing data, the practice is by no means universal. A study co-funded by the European Commission in 2009 examined the research habits of over 1200 researchers and found that the primary barriers for research data sharing included potential legal issues, misuse of data, incompatible data types, insufficient technical infrastructure and/or financial resources (Kuipers & van der Hoeven, 2009). The reasons not to share are complicated but can be broadly categorized as either technical or cultural (Aalbersberg, Dunham, & Koers, 2013; Borgman, 2012; Reilly et al., 2011; Smith et al., 2011; Thessen & Patterson, 2011; Wallis et al., 2013). Thessen and Patterson (2011) contend that the cultural barriers will be the most difficult to overcome; some life science disciplines such as molecular biology, embrace data sharing, while others such as zoology are just starting to do so.

## BARRIERS TO DATA SHARING

Data sharing among the sciences does not have an all-inclusive, uniform or over-arching data culture, a function of both values and of sheer quantities and types of data produced. Thessen and Patterson (2011, p. 19) have described "data culture" as referring to "the explicit and implicit data practices and expectations that determine the destiny of data. It relates to the social conventions of acquisition, curation, preservation, sharing, and reuse of data". Each discipline has its own "data culture", for example, field biologists collect their data in lab or note books "as a prelude to a narrative explanation of observations to the molecular biologist whose data are born digital in near terabyte quantities and are widely shared through global data repositories" (Thessen & Patterson, 2011, p. 19). Other data characteristics can also have an impact on researchers' willingness to share. Data from "big science", or large, collaborative, often long-term projects, are intended to be shared among large teams and they are typically more uniform and therefore more easily transferable. "Small science" data, or the "long tail" of science is usually generated by smaller research teams in more idiosyncratic formats that are not easily transferable beyond the team

Download English Version:

https://daneshyari.com/en/article/358322

Download Persian Version:

https://daneshyari.com/article/358322

Daneshyari.com