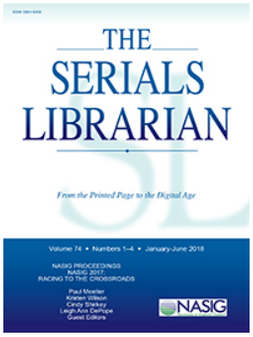

The Serials Librarian

From the Printed Page to the Digital Age

ISSN: 0361-526X (Print) 1541-1095 (Online) Journal homepage: http://www.tandfonline.com/loi/wser20

Data Stories: Using Narratives to Reflect on a Data Purchase Pilot Program

Anita K. Foster & Gene R. Springs

To cite this article: Anita K. Foster & Gene R. Springs (2018) Data Stories: Using Narratives to Reflect on a Data Purchase Pilot Program, The Serials Librarian, 74:1-4, 128-132, DOI: 10.1080/0361526X.2018.1427991

To link to this article: https://doi.org/10.1080/0361526X.2018.1427991

THE SERIALS LIBRARIAN
2018, VOL. 74, NOS. 1–4, 128–132
https://doi.org/10.1080/0361526X.2018.1427991

Data Stories: Using Narratives to Reflect on a Data Purchase Pilot Program

Anita K. Foster and Gene R. Springs

Presenters

ABSTRACT

Acquiring research data by university libraries is a difficult process to manage. It involves difficulties in identifying data available for purchase, funding data purchases, and getting acceptable licensing terms. During fiscal year 2017, the administration of The Ohio State University Libraries issued a strategic focus collection area mandate to acquire research data sets that could be used by the entire university population for various research needs. This article describes the institutional background that led to this strategy mandate, how the Libraries responded, and a pilot purchase program for research data sets developed through a collaboration between the Libraries' Collections Strategist and Electronic Resources Officer. It will also discuss potential workflows for collection development and acquisitions for potential future purchases of research data sets.

KEYWORDS

Data acquisition; data licensing; research data

Introduction

In response to the increasing needs regarding research data at The Ohio State University (OSU), the administration of The Ohio State University Libraries (OSUL) made purchasing data a strategic collection area in fiscal year 2017. Although data have been purchased in a variety of formats and across numerous disciplines for decades, a focused effort to identify, purchase, and make accessible data was a new challenge for OSUL. This article details the background that led to the administrative mandate and the process the OSUL Collections Strategist and Electronic Resources Officer undertook to acquire relevant data for the maximum benefit of the OSU research community.

Research data at The Ohio State University

At OSU, the Discovery Theme is a university-wide initiative that sets institutional priorities for both academic and research investment. Discovery Themes are linked to the land grant mission of OSU and are geared toward finding durable solutions to the most compelling issues facing the world.[1] Following a proposal process, Translational Data Analytics was announced as a Discovery Theme in 2014, and a $500 million investment was made to develop academic programs, spur research, and hire new tenure-track faculty around four core areas: Humanities and the Arts, Health and Wellness, Food Production and Security, and Energy and Environment.[2] The interdisciplinary nature of the Discovery Theme's core areas cut across the administrative structure of OSU, placing an emphasis on collaboration and partnership. OSUL, as a unit that already supports academic and research needs across the entire university, was well positioned to develop and deploy services in support of Translational Data Analytics.

CONTACT Gene R. Springs springs.14@osu.edu 350B Thompson Library, 1858 Neil Avenue Columbus, OH 43210, USA.

OSUL identified the need to support collaborative, interdisciplinary research by graduate students and faculty in 2012 and began planning for a library facility that could provide diverse research support services, including those around data. In 2013, OSUL advertised a cluster hire initiative that included three data-focused positions to be based in the planned Research Commons: Data Management Services Librarian, Digital Humanities Librarian, and Geographic Information Systems (GIS) Librarian. Although initially only the Data Management Services Librarian search was successful in 2014, a Digital Humanities Librarian did join OSUL in 2016, and GIS research support has been provided by the manager of the Research Commons. Support services from the Research Commons, including workshops and consultations on data management plans, data ethics, and GIS software, were deployed before the facility itself opened in the spring of 2016, establishing OSUL as a central hub on campus for data-related research support.

In 2016, after the arrival of a new Vice Provost and Director of Libraries, the Research Data in the Library task force was charged with examining the research data services needs of the OSU campus and making recommendations for potential future support services OSUL could develop and deploy. Although there had been previous discussions at OSUL about data services writ large,[3] particularly through a forum dedicated to the topic in fall of 2014, the task force would provide a more comprehensive environmental scan of campus needs. The task force, chaired by the Data Management Services Librarian, had wide representation from various OSUL departments, including two subject librarians, the Libraries' GIS specialist, the Digital Humanities Librarian, the University Archivist, the Head of Publishing and Repository Services, the Electronic Resources Officer, the Collections Strategist, and a librarian from the OSU Prior Health Sciences Library. The task force met over the course of one summer and the following academic year, revising initial recommendations. The end result was the creation of a Data Working Group to carry forward data-related work at the OSUL.

The momentum for data needs for OSU researchers escalated with the opening of the Research Commons facility in spring 2016. Having a place for collaborative, interdisciplinary research naturally led to questions from users about available research data. In July 2016, the Collections Strategist received instructions from OSUL administration that purchasing data to support the research needs of the OSU community was one of three strategic collection areas for the new fiscal year, alongside streaming media and business resources. The challenge of acquiring research data was clear: research needs can be very specific, so how might OSUL go about purchasing data that can benefit the most researchers at OSU?

Given little criteria beyond campus-wide access, one-time payments for multi-year access, and no specific budget, the Collections Strategist and Electronic Resources Officer used several initial strategies to develop a list of datasets and data resources to explore. First, a call for suggestions went out to the subject librarians and collection managers to recommend resources their constituent communities may have inquired about in the past. Next, the Collections Strategist followed up on a few data resource requests that came directly from faculty to him. Finally, the Electronic Resources Officer re-examined data resources that had been explored in the past but not pursued for a variety of reasons. These strategies resulted in the exploration of eleven total data resources, of which seven were purchased; of the seven purchased, four were typical resources hosted on a vendor platform, while three required local hosting, which complicated both the acquisitions process and the potential for access for the OSU community (see Figure 1).

Licensing considerations for research data sets

Research data sets exist in a variety of formats and are available either through traditional web-based interfaces or by hosting the data locally on networked servers or other platforms. Data sets are also purchased from vendors familiar with working with libraries and those who are newer to the world of academia. Each data set purchased in the pilot program had its own licensing quirks, but some common characteristics emerged as well.

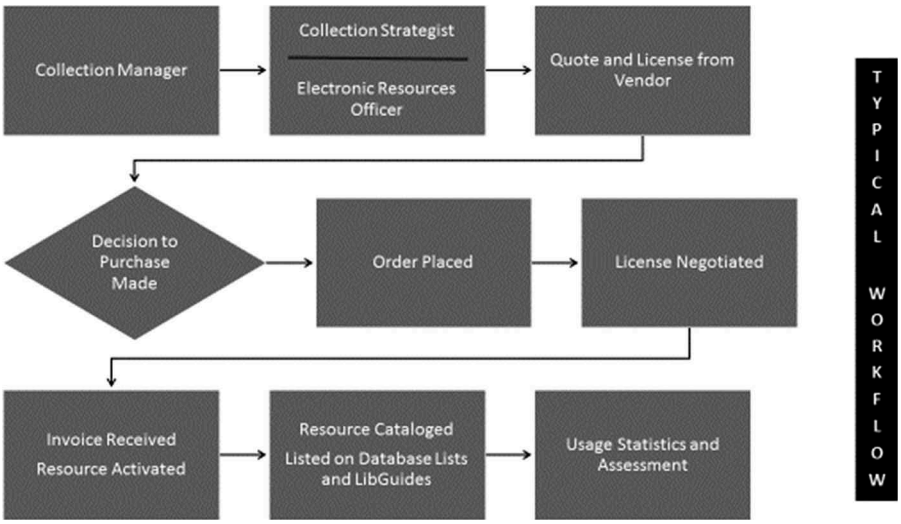

Figure 1. Typical workflow for acquiring electronic resources.

Standard issues around jurisdiction and indemnification are seen in licenses for research data, but there are other areas that arose more often with these types of purchases for typical electronic resources acquisitions. Common characteristics in licensing research data sets include auditing language and post-cancellation requirements for the library. Additionally, there are business-specific terms to these licenses, such as more detailed language around confidentiality and possible litigation.

Auditing language is a difficult licensing area for libraries. Auditing of financial books and other systems is common in industry but is less frequently mentioned in academia. Auditing of computer networks for files that are no longer authorized files is also uncommon. Terms for both types of auditing were seen during this project. While institutionally not prohibited, auditing terms for the systems were not viewed favorably in the Libraries. Each license with auditing clauses was handled differently, depending on vendor response. Agreed-on auditing terms varied from the Libraries certifying certain factors had been accomplished to a completed form to standard auditing processes. But in each case, the Libraries retained rights over access and accountability of the vendor representatives doing the work.

Another common theme was language detailing what would happen when the relationship between the vendor and the Libraries ended, a form of post-cancellation rights. Many licenses only invoked auditing rights at that point in the process, not earlier. Vendors wanted to ensure that all copies of their data had been deleted or returned and were no longer in use once the agreement ended. This requirement poses some difficulties for a library. At OSU, networks are mostly decentralized and managed by individual colleges and departments. This concept had to be described and defined for some vendors; in particular, those who worked more often with business entities and not academia. The Libraries could agree to terms that governed actions only for its own servers, not the servers controlled by other departments. In order to finalize some licenses, OSUL and the vendors had to compromise and agree to specific language that met the vendor's need and allowed OSUL to fulfill any known obligation.

Language less commonly seen in standard electronic resource licensing encountered in this project included terms around export compliance/control and whether the library could inform patrons about the availability of the research data sets purchased. As OSUL was purchasing research data sets that could be utilized by researchers in disciplines across OSU, it was imperative that normal communication methods could be used to inform the community of their availability. Clarity

and terminology regarding the inclusion of information in the catalog, lists of databases, and LibGuides was added to at least two of the licenses for acquired data.

One of the research data products had an export compliance clause in the license that almost ended the pursuit of the resource. Export compliance laws and terms govern technology and other commodity transfers outside the United States. As a large university, OSU has a substantial international community of students and researchers, many of whom travel to countries that may be on export control embargo lists or blocked for access for other reasons. While seeing this clause in a license for locally hosted data would potentially be manageable for a library, the ability to police activity in a web-based resource is not possible with current authentication systems. For example, while EZProxy has ways to restrict access for specific regions, it cannot do so for individual resources, and so having one system with this requirement was unworkable. The Electronic Resources Officer talked with the campus Export Control Office to learn more about the university regulations and what, if any, flexibility OSUL had for managing such terms in resource licenses. It was clear that this type of export compliance term in this license was untenable and would need to be removed or radically changed. Ultimately, OSUL and the vendor compromised on the export compliance language and the license was completed.

While not all institutions have the funding or ability to pay upfront for multi-year deals, OSUL has been doing so for a number of years. There are pros and cons to doing so, but it was one of the few criteria given to the project members for this pilot. In terms of licensing, all but one of the contacted vendors was willing to use this model. The license agreements for all purchased data sets included language that detailed how and when new data files would be available and delivered. For the locally hosted data sets, data format was also frequently included in the agreement.

Data set collecting and process change

One aspect of this strategic focus collecting pilot was an intent to develop a collection development process for possible future purchases of research data sets. In evaluating the project, the Collection Strategist and the Electronic Resources Officer identified what worked and what did not during the

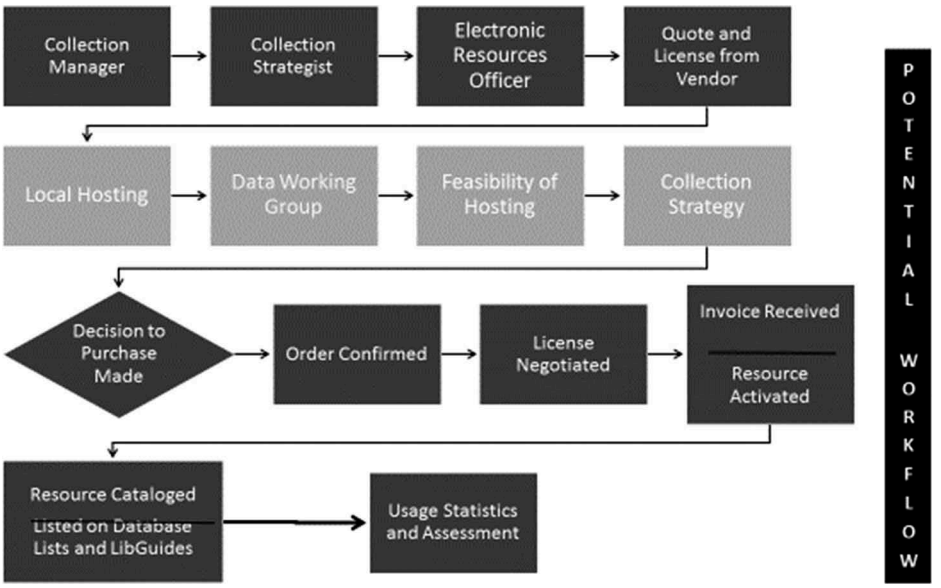

Figure 2. Revised collection development and acquisitions workflow for data.

year. In the end, they developed additional criteria to be applied when such a data purchase is being considered in the future. The criteria are:

(1) Selection decision based on the appropriateness of data content and potential user population
(2) Data purchased by OSUL is for use by the whole university
(3) Is the data accessible via the web or is local hosting required?
(4) For web-based data, follow standard workflow for electronic resource acquisitions
(5) For locally hosted data, purchase requests will be discussed by members of the Data Working Group focused on the type of data, the feasibility of hosting, and options for user access
(6) Funding source and payment type (e.g., multi-year pay in advance)

Slight modifications to the collection development and acquisitions workflows will be made to accommodate buying research data sets. The overall workflow will have little change (see Figure 2).

Conclusion

While the start of the research data set collection pilot was often fraught with uncertainties, a structured method was developed that is workable and flexible in a way that can adapt to changes in research needs and availability of large data sets. The strategic collection priority for purchasing data to support the escalation of data initiatives from stakeholders on the OSU campus allowed OSUL to engage with the acquisition of data sets and eventually access workflows in real time. A major consideration in closing the loop on this pilot project in the future will be how assessment metrics are developed, specifically for the locally hosted data.

Notes

1. OSU Discovery Themes. "Guiding Principles," https://discovery.osu.edu/about/guiding-principles.html (accessed July 23, 2017).
2. "Development of a Singular Presence in Data Analytics for The Ohio State University," https://discovery.osu.edu/assets/documents/DataAnalyticsExecutive%20Summary.pdf (accessed July 23, 2017); OSU Discovery Themes, "Theme Areas" https://discovery_test.org.ohio-state.edu/assets/documents/DataAnalyticsExecutive%20Summary.pdf (accessed July 23, 2017).
3. Juleah Swanson and Amanda Rinehart, "Data in Context: Using Case Studies to Generate a Common Understanding of Data in Academic Libraries," *Journal of Academic Librarianship* 42, no. 1 (2016): 97–101.

Notes on contributors

Anita K. Foster is Electronic Resources Officer at The Ohio State University Libraries, Columbus, Ohio.

Gene R. Springs is Collections Strategist at The Ohio State University Libraries, Columbus, Ohio.